



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA INFORMAČNÍCH TECHNOLOGIÍ**

FACULTY OF INFORMATION TECHNOLOGY

**ÚSTAV INTELIGENTNÍCH SYSTÉMŮ**

DEPARTMENT OF INTELLIGENT SYSTEMS

**APLIKACE PRO GENEROVÁNÍ IDENTIKITU NA  
ZÁKLADĚ TEXTOVÉHO POPISU V ČEŠTINĚ**

APPLICATION FOR GENERATING AN IDENTITY BASED ON A TEXT DESCRIPTION IN CZECH

**BAKALÁŘSKÁ PRÁCE**

BACHELOR'S THESIS

**AUTOR PRÁCE**

AUTHOR

**JAN FURIK**

**VEDOUcí PRÁCE**

SUPERVISOR

**Ing. TOMÁŠ GOLDMANN, Ph.D.**

BRNO 2024

## Zadání bakalářské práce



164347

Ústav: Ústav inteligentních systémů (UITS)  
Student: **Furik Jan**  
Program: Informační technologie  
Název: **Aplikace pro generování identikitu na základě textového popisu v češtině**  
Kategorie: Umělá inteligence  
Akademický rok: 2024/25

### Zadání:

1. Seznamte se s problematikou generování snímků obličejů pomocí neuronových sítí. Prozkoumejte dostupné technologie a knihovny pro generování snímků obličeje na základě textového popisu.
2. Nastudujte proces tvorby identikitů. Zaměřte se na standardní postupy používané v kriminalistice a antropologii.
3. Navrhněte aplikaci, která dokáže na základě popisu vygenerovat identikit.
4. Aplikaci implementujte v programovacím jazyce Python.
5. Proveďte testování aplikace. Dále proveďte praktické testování s kontrolní skupinou uživatelů.

### Literatura:

- RYAN, Michael. Facial Composite Generation from Natural Language Text Descriptions using Stacked Generative Adversarial Networks.
- JALAN, Harsh Jaykumar, et al. Suspect face generation. In: *2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA)*. IEEE, 2020. p. 73-78.

Při obhajobě semestrální části projektu je požadováno:  
Body 1 a 2.

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Goldmann Tomáš, Ing., Ph.D.**  
Vedoucí ústavu: Kočí Radek, Ing., Ph.D.  
Datum zadání: 1.11.2024  
Termín pro odevzdání: 14.5.2025  
Datum schválení: 31.10.2024

## Abstrakt

Tato bakalářská práce se zaměřuje na generování identikitů na základě textového popisu pomocí generativních neuronových sítí. Cílem je vyvinout aplikaci v jazyce Python, která umožní vytvoření portrétní fotografie osoby na základě slovního popisu v češtině. Práce se zaměřuje na analýzu metod portrétní identifikace, využití text-to-image technologií a implementaci systému s možností manuální úpravy výsledného obrazu. Navržené řešení využívá pokročilé techniky strojového učení a přispívá k propojení forenzní praxe s moderním vývojem v oblasti umělé inteligence.

## Abstract

This bachelor thesis focuses on generating identikit based on textual description using generative neural networks. The goal is to develop a Python application that allows the creation of a portrait of a person based on a verbal description in Czech. The work focuses on the analysis of portrait identification methods, the use of text-to-image technologies and the implementation of a system with the possibility of manual editing of the resulting image. The proposed solution uses advanced machine learning techniques and contributes to the connection of forensic practice with modern developments in the field of artificial intelligence.

## Klíčová slova

kriminalistika, portrétní identifikace, identikit, neuronové sítě, generování syntetických obličejů

## Keywords

criminalistics, portrait identification, identikit, neural networks, synthetic face generation

## Citace

FURIK, Jan. *Aplikace pro generování identikitu na základě textového popisu v češtině*. Bakalářská práce. Vedoucí práce Ing. Tomáš Goldmann, Ph.D.. Brno: Vysoké učení technické v Brně, Fakulta informačních technologií, 2024.

# Aplikace pro generování identikitu na základě textového popisu v češtině

## Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením Ing. Tomáše Goldmanna, Ph.D. Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

.....

Jan Furik  
12. května 2025

## Poděkování

Rád bych poděkoval panu Ing. Tomáši Goldmannovi, Ph.D., za odborné vedení a cenné rady při vytváření této bakalářské práce. Dále bych chtěl poděkovat za podporu své rodině, zejména svému otci Mgr. Miroslavu Furikovi, důstojníkovi Policie ČR, který mi poskytl cenné podněty a inspiraci při zpracování odborných částí této práce.

# Obsah

<b>1</b>	<b>Úvod</b>	<b>4</b>
<b>2</b>	<b>Portrétní identifikace</b>	<b>5</b>
2.1	Význam a definice portrétní identifikace . . . . .	5
2.2	Druhy popisu osob . . . . .	6
2.3	Metody vytváření identikitu . . . . .	9
2.4	Aplikace pro vytváření identikitu . . . . .	10
2.5	Kognitivní funkce ovlivňující tvorbu identikitu . . . . .	12
<b>3</b>	<b>Generování snímků obličeje pomocí neuronových sítí</b>	<b>14</b>
3.1	Úvod do neuronových sítí . . . . .	14
3.2	Model neuronu . . . . .	14
3.3	Generativní soupeřící síť . . . . .	17
3.4	Variační autoenkodér . . . . .	18
3.5	Difúzní modely . . . . .	19
3.6	Autoregresivní model . . . . .	21
3.7	Text-to-image modely . . . . .	21
3.8	ControlNet . . . . .	23
3.9	Image inpainting . . . . .	24
<b>4</b>	<b>Návrh řešení</b>	<b>25</b>
4.1	Specifikace požadavků . . . . .	25
4.2	Výběr technologií . . . . .	25
4.3	Použité knihovny . . . . .	26
4.4	Návrh aplikace . . . . .	27
4.5	Návrh masek . . . . .	29
4.6	Aplikační rozhraní . . . . .	30
<b>5</b>	<b>Implementace</b>	<b>31</b>
5.1	Generování obličejů . . . . .	31
5.2	Editace obličejů . . . . .	32
5.3	Manuální úprava obličeje . . . . .	33
5.4	Implementace frontendu . . . . .	35
5.5	Výpočetní náročnost . . . . .	36
<b>6</b>	<b>Testování</b>	<b>37</b>
6.1	Experiment 1 – Popis i generování jednou osobou . . . . .	37
6.2	Experiment 2 – Popis svědkem, zadávání úředníkem . . . . .	38

6.3	Experiment 3 – Popis z paměti . . . . .	40
6.4	Testování uživatelského rozhraní . . . . .	42
6.5	Výsledky testování . . . . .	43
<b>7</b>	<b>Závěr</b>	<b>44</b>
	<b>Literatura</b>	<b>45</b>
<b>A</b>	<b>Obsah paměťového média</b>	<b>48</b>
<b>B</b>	<b>Sekvenční diagram</b>	<b>49</b>

# Seznam obrázků

2.1	Aplikace FACES [19]. . . . .	11
3.1	Umělý neuron [6]. . . . .	15
3.2	Architektura trénování modelu GAN [24]. . . . .	18
3.3	Generování pomocí ControlNet Canny. [28]. . . . .	23
4.1	Architektura aplikace. . . . .	26
4.2	Ukázka uživatelského rozhraní aplikace. . . . .	29
4.3	Masky reprezentující tvar obličeje. . . . .	30
5.1	Vygenerované snímky obličeje. . . . .	32
5.2	Upravený snímek obličeje. . . . .	33
6.1	Reálná fotografie [18] a vygenerovaný identikit Leoše Mareše. . . . .	38
6.2	Reálná fotografie [17] a vygenerovaný identikit Scarlett Johansson. . . . .	39
6.3	Reálná fotografie [11] a vygenerovaný identikit Davida Pastrňáka. . . . .	39
6.4	Reálná fotografie [25] a vygenerovaný identikit Daniela Radcliffa. . . . .	40
6.5	Reálná fotografie [31] a vygenerovaný identikit Lucie Bílé. . . . .	41
6.6	Reálná fotografie [4] a vygenerovaný identikit Zandayi. . . . .	41
6.7	Reálná fotografie [15] a vygenerovaný identikit Vicy Kerekes. . . . .	42
B.1	Sekvenční diagram aplikace. . . . .	49

# Kapitola 1

## Úvod

Generování identikitů na základě slovního popisu představuje zajímavé téma, které se nachází na pomezí kriminalistiky, psychologie a moderních technologií využívajících neuronové sítě. V oblasti portrétní identifikace se tradičně využívá metoda skládaného portréту, jehož výsledná kvalita i přesnost závisí na schopnosti svědka popsat vzhled osoby a na dovednosti operátora efektivně interpretovat popis. Vývoj v oblasti umělé inteligence, především v generativních neuronových sítích, přináší nové možnosti, které by mohly tento proces automatizovat a zpřesnit.

Cílem této bakalářské práce je prozkoumat možnosti generování obličejů pomocí neuronových sítí a dostupných technologií pro tvorbu obrázků na základě textového popisu a nastudovat proces tvorby identikitů s ohledem na kriminalistické a antropologické postupy. Následně navrhnout a implementovat aplikaci v jazyce Python, která umožní generovat identikit z českého textového popisu.

První část práce se věnuje teoretickému základu portrétní identifikace, popisu osob a rozdílu mezi laickým a úředním popisem, včetně využívané terminologie v kriminalistické praxi. Následuje přehled metod pro sestavení portréту a představení některých aplikací využívaných pro tvorbu identikitu. Práce se též zaměřuje na psychologické faktory ovlivňující vnímání a rekonstrukci paměťové stopy, včetně experimentů provedených v rámci studie zaměřené na rozpoznávání tváří.

V další části jsou představeny neuronové sítě, jejich architektura a modely vhodné pro generování obrazových dat. Podrobně jsou popsány moderní text-to-image přístupy, které tvoří základ navrženého systému. Pozornost je též věnována metodám umožňujícím přesnější ovlivnění výsledného obrazu za pomoci doplňujících vstupů, jako je například úprava konkrétních oblastí nebo řízení struktury a kompozice generované scény.

Praktická část práce se věnuje návrhu a implementaci aplikace v jazyce Python, včetně uživatelského rozhraní a podpory manuální úpravy výsledného obrázku. Funkčnost systému je ověřena pomocí série experimentů a testována i v rámci kontrolní skupiny uživatelů, jejichž zpětná vazba přispěla k vyhodnocení použitelnosti řešení.

Aplikace využívá pokročilé techniky strojového učení, jako jsou difúzní modely a rozšiřující mechanismy pro řízení generování, díky nimž je možné vytvářet vizuální podobu osoby na základě textového popisu, včetně možností cílené editace obrazu a manuální manipulace s jeho prvky.

Tato práce tak propojuje kriminalistickou praxi s moderním vývojem v oblasti umělé inteligence a nabízí praktickou ukázkou využití neuronových sítí v oblasti forenzní vizualizace.

## Kapitola 2

# Portrétní identifikace

Portrétní identifikace představuje kriminalistickou metodu zjišťování totožnosti osob na základě jejich vnějších znaků. Vychází z předpokladu, že každý člověk má jedinečnou kombinaci morfologických a metrických charakteristik, které lze systematicky popsat a využít k identifikaci. Tato metoda se dělí na úřední a laický popis, přičemž úřední popis je prováděn odborníkem podle přesně stanovených pravidel. Zaměřuje se na znaky relativně neměnné (např. tvar lebky, obličej, uši) i proměnlivé (např. vrásky, vlasy, vousy). Portrétní identifikace je důležitým nástrojem kriminalistické praxe, zvláště v případech, kdy nejsou k dispozici jiné identifikační metody [23].

### 2.1 Význam a definice portrétní identifikace

*„Portrétní identifikace je vědní obor kriminalistické techniky zabývající se zkoumáním zákonitostí vnějších znaků osoby a rozpracováním metod a prostředků na zjišťování totožnosti osob podle jejich vnějších znaků pro potřeby kriminalisticko-bezpečnostní praxe“ [29, s. 62].*

Tato část je částečně převzata z [29]. Pojem portrétní identifikace lze vymezit jako postup identifikace osob na základě relativně stálých fyziologických znaků. Jedná se o nejstarší identifikační metodu individuální identifikace osob, jejíž základy položil v roce 1879 Alphonse Bertillon. Jeho identifikační metoda popisovala 11 různých tělesných rozměrů osob, které byly zapsány na jedné kartě. Kartu bylo možné na základě jednoduchého systému rychle vyhledat v kartotéce.

Z antropologického hlediska je každý člověk charakterizován souborem relativně stálých morfologických a metrických znaků, jejichž komplex je označován jako vnější znaky, které se u žádné jiné osoby nevyskytují stejně.

Vnější znaky osob lze dělit na:

- **statické** (anatomické), které jsou dány stavbou těla,
- **dynamické** (funkční), které lze hodnotit, pokud je osoba v pohybu (lokomoce, gestikulace, běh, chůze, skok, způsob mluvy).

Při kriminalisticko-technické identifikaci lze kromě tělesných znaků vyjádřitelných metrickými údaji popsat osobu také řadou morfologických charakteristik, které měřením vyjádřit nelze. Jedná se například o vyjádření obrysu obličej, tvar rtů, nosu, záhybu horního víčka apod. Popis osoby by měl obsahovat co nejvýstižnější morfologické znaky. Vnější obličejové znaky lze rozdělit do několika kategorií podle různých hledisek.

Znaky z hlediska fyzické charakteristiky:

- **morfologické** – kvalitativní (tvar rtů, tvar oční štěrbin, průběh hřbetu nosu...),
- **metrické** – kvantitativní (výška obličeje, vzdálenost očních štěrbin...).

Znaky z hlediska výpovědní hodnoty:

- **Obecné** (skupinové, typové) – jedná se o znaky, které se vyskytují běžně v populaci. Zde se hodnotí charakteristika tvaru a relativní míra vyhnutí.
- **Specifické** (individuální, zvláštní) – v populaci jsou velmi málo frekventované. Dělí se na znaky vrozené (rozštěp patra atd.) a znaky získané (jizvy, vrásky, tetování, atd.).

Znaky z hlediska relativní proměnlivosti:

- **S věkem zanedbatelně proměnlivé** – v průběhu vývoje nedochází k významnějším změnám (reliéf ušního boltce s výjimkou lalůčku).
- **S věkem proměnlivé** – zde je patrný vliv stárnutí (ztráta zubů, pokles víček, vrásky).

## 2.2 Druhy popisu osob

Při popisu osoby rozlišujeme popis úřední a laický, přičemž úřední se řídí odbornými pravidly a laický je subjektivní a nesystematický.

### 2.2.1 Popis úřední

Tato část je inspirována [29, 20]. Popis úřední provádí úřední osoba, zpravidla speciálně vyškolený specialista. Úřední popis vyžaduje soustředěné pozorování a vnímání charakteristických znaků, se znalostí terminologie, k provedení co nejpřesnějšího popisu jednotlivých znaků. Výhodou je, že osoba, která provádí tento popis, má možnost popisovanou osobu pozorovat zblízka za denního světla a ze všech stran. Má možnost s osobou komunikovat a sledovat její způsob mluvy, gestikulaci a chůzi.

Z důvodu snížení vzniku subjektivních rozdílů při popisování osob vznikl jednotný systém, který popisuje jednotlivé tělesné atributy: tělesná výška, hmotnost, postava, zdánlivé stáří, tvar lebky a obličeje, vlasy, čelo, obočí, uši, oči, nos, vousy, ústa, rty, zuby a brada.

#### Tělesná výška

Jedná se o nejobecnější identifikační znak, který podléhá v průběhu věku minimálním změnám. Údaj o výšce osoby se zjišťuje jejím měřením ve stoje. Pokud nelze osobu změřit bez obuvi, odečte se po provedeném měření výška podrážky. U mrtvol se tělesná výška měří ve vodorovné poloze a vždy vedle těla.

Rozdělení výšky osob: velmi malá (do 150 cm), malá (do 160 cm), střední (do 170 cm), velká (do 180 cm), velmi velká (do 190 cm) a obrovitá (nad 200 cm).

#### Tělesná hmotnost

Tento údaj se zjišťuje přesným vážením, ideálně pouze ve spodním prádle. V odůvodněných případech se od zjištěné hmotnosti odečte váha oblečení. Ve výjimečných situacích lze tělesnou hmotnost odhadnout. Hmotnost člověka se může během poměrně krátké doby výrazně

změnit. Jedná se o údaj, který ve spojení se zjištěnou výškou vytvoří relevantní představu o postavě.

### **Postava**

Pod tímto pojmem si lze představit celkový vzhled osoby, kterou popisujeme. Jde o poměr mezi výškou, šířkou jednotlivých částí těla a proporcemi břicha a hrudníku, přičemž typ postavy určujeme na základě těchto poměrů a rozměrů jednotlivých částí těla. Typy postavy dělíme na slabou, střední, zavalitou a silnou.

### **Zdánlivé stáří**

Uvádí se pouze v případech, kdy je skutečný fyzický věk osoby v nesouladu s jejím vzhledem.

### **Tvar hlavy**

Je popisován zepředu. Jedná se o jeden z nejdůležitějších identifikačních znaků jedince. Rozlišujeme tyto tvary: kulatá, čtvercová, oválná, obdélníková, pyramidová, vejčitá, kosočtverečná, dvojevypuklá a nepravidelná.

### **Tvar lebky**

Jedná se o tvar při pohledu z profilu. K jejímu popisu lze využít tyto pojmy: vysoká, nízká, špičatá, tupá, vejčitá, plochá, dopředu vyklenutá, dozadu vyklenutá, s čelním hrbolem, s temenním hrbolem a nepravidelná.

### **Obličej**

Jeho popis je ohraničen vlasy, spánkovými a lícními kostmi, dolní čelistí a bradou. Při popisu obličeje se uvádí jeho tvar, barva, plnost a případné zvláštnosti. Při popisu uvádíme také umístění vrásek, dolíčků, kosmetických vad, četnost pih apod. U tvaru obličeje (koresponduje s tvarem hlavy) rozlišujeme: kulatý, oválný, obdélníkový, pyramidový, vejčitý a kosočtvercový. Při popisu barvy obličeje se uvádí pokud možno jeho přirozená barva (nikoli získaná např. opálením). Rozeznáváme tyto barvy: bledá, růžová, žlutá, snědá, nebo zvláštní zbarvení (vyjadřuje se přirovnáním, např. cihlově červená atd.).

### **Vlasy**

Při jejich popisu uvádíme barvu, tvar, hojnost (hustotu), střih a další případné zvláštnosti. Barvu určujeme podle vzorkovnice v rozsahu 13 barev. Mezi základní tvary vlasů patří: rovné, vlnité, kadeřavé či kudrnaté. Dle hojnosti rozeznáváme vlasy: husté, řídké, částečná holohlavost (temenní, čelní a postranní) či úplná holohlavost. U střihu a účesu uvádíme například: sčesání vzad, sčesání vpřed, sčesání do stran (uvést umístění pěšinky).

### **Čelo**

Popis obsahuje výšku, šířku, sklon, klenutí a vrásky. Výška čela se uvádí v poměru k délce hlavy. Je-li výška čela větší než třetina hlavy, mluvíme o čele vysokém. Pokud má výšku menší mluvíme o čele nízkém. Hranice vlasů není pro délku čela určující. Šířka čela se uvádí v poměru k šířce obličeje. Pokud je čelo širší než obličej, hovoříme o čele širokém. Pokud je šířka čela menší než šířka obličeje, hovoříme o čele úzkém. U čela dále určujeme jeho sklon, který se určuje podle úhlu, který svírá myšlená přímka protínající kořen nosu při pohledu z profilu se sklonem čela. Na základě zjištěného úhlu rozlišujeme čelo kolmé, šikmé (dozadu) nebo vysedlé (dopředu). Jako zvláštní znak se vyskytují čelní hrboly, silně vyvinuté nadočnicové oblouky, či čelo zvlnělé. Na čele lze dle umístění popisovat vrásky kolmé, příčné, nadočnicové a nad kořenem nosu. Popisujeme jejich počet, tvar, velikost a hloubku.

## **Obočí**

Popisujeme tvar, sklon, velikost, vzdálenost a barvu. Dále podle tvaru rozeznáváme obočí přímé, obloukovité, lomené a vlnité. Obočí může být skloněno šikmo dovnitř ke kořenu nosu, nebo naopak skloněno ke spánkům. Dle velikosti dělíme obočí na krátké, dlouhé, široké a úzké. Obočí může být daleko od sebe, blízko u sebe, nebo srostlé. Dále se uvádí hustota či řídkost obočí, včetně nepravidelného růstu (např. holé ostrůvky). Barvu uvádíme pouze, pokud je odlišná od barvy vlasů. Ke zvláštnostem patří obočí různobarevné, s barevnými skvrnami, velmi husté či vyholené.

## **Oči**

Při popisu očí uvádíme jejich barvu, tvar oční štěrbinu a případné zvláštnosti (např. různá barva duhovek). Barvu popisujeme ze vzdálenosti cca 30 cm při denním světle. U oka rozeznáváme tyto barvy: modrá, šedomodrá, šedozelená, modrozelená, šedohnědá, zelenohnědá, světlehnědá a tmavohnědá. U oční štěrbinu se uvádí, zda osa očních koutků je vodorovná, šikmá dovnitř nebo šikmá vně. Zvláštností mohou být oči vystouplé, vpadlé, šilhavé, se zvláštní barvou bělma, skleněné nebo s viditelnou vadou. Dále u očí uvádíme nápadně malou či velkou velikost a zda jsou od sebe široce rozložené, nebo blízko rozložené.

## **Ucho**

Tvar ucha je v průběhu života téměř neměnné. Zpravidla se popisuje pravé ucho, levé se popisuje pouze v případě zřetelných odchylek. Rozlišujeme tyto tvary: oválné, kulaté, trojhranné a čtyřhranné. U ucha dále rozlišujeme způsob jeho usazení, zda je rovné nebo šikmé (s uvedením směru vychýlení). Ušní lalůček může mít tvar okrouhlý, pravoúhlý nebo klínový. Lalůček může být dále volný nebo přirostlý. Mezi zvláštnosti patří odstávající nebo velmi přiléhavé boltce, zvláštní sklon, výrůstky, neobvyklý tvar, propíchnutý lalůček, nízké, vysoké nebo nepravidelné uložení uší.

## **Nos**

Určujeme jeho šířku, výšku, profil, hřbet, základnu, kořen a špičku nosu. Pokud šířka chřípí značně přesahuje šířku nosu, mluvíme o nose širokém a naopak. Pokud je délka nosu delší, než je jedna třetina délky obličeje, mluvíme o nose dlouhém a v opačném případě o krátkém. Hřbet nosu může být vyhloubený, přímý, vypouklý nebo orlí. Základna nosu může být dále vodorovná, zdvižená nahoru či skloněná dolů. Špička nosu může být úzká, široká nebo rozpůlená. U kořene nosu rozeznáváme, zda je plochý, hluboký, úzký nebo široký. Dále mohou být zaznamenány různé zvláštnosti, například nos zmáčknutý, zjizvený nebo zduřelý nos.

## **Vousy**

Pro identifikaci nemají větší význam, neboť je jejich nositel může upravit nebo oholit. U vousů se rozeznává jejich barva, hustota, vzrůst, tvar a stříh.

## **Ústa a rty**

Popisujeme celkový vzhled a tvar. Velikost se popisuje v porovnání s šířkou nosu. Může se jednat o ústa malá, střední, nebo velká. Výška je určována šířkou rtů.

## **Zuby**

Mají důležitou identifikační hodnotu s charakteristickými znaky. Chrup má standardně 32 zubů umístěných v horním a dolním zubním oblouku čelistí. Důležitým identifikačním znakem jsou veškeré lékařské nebo dentistické zásahy. Při popisu uvádíme celkový stav, umělý chrup, můstek a případné zvláštnosti.

## **Brada**

Popisujeme tvar zředu a ze strany. Mezi zvláštnosti patří důlek, rýha, dvojitá brada apod.

### **Zvláštní znamení a tetování**

Jedná se o jizvy, bradavice, mateřská znaménka, dolíčky po neštovicích apod. Při popisu jizev uvádíme velikost, tvar a polohu.

### **2.2.2 Laický popis**

Tato část textu je založena na zdrojích [29, 23]. Popis pachatele je získáván odborně vyškoleným policistou na základě výslechu osoby (svědka či poškozeného). Výsledkem se policista snaží získat přesný popis vnějších znaků, které následně převede do terminologie úředního popisu. Kvalitu získaného popisu mohou ovlivnit objektivní či subjektivní okolnosti.

**Objektivní okolnosti** jsou vnější vlivy, které existují nezávisle na vůli a vnímání osob. Jedná se například o vzdálenost, viditelnost, denní dobu a směr pozorování.

**Subjektivní okolnosti** jsou vlivy, které jsou dány vlastnostmi a schopnostmi vyslychaného. Především se jedná o schopnost vnímanou událost udržet v paměti a později ji přesně reprodukovat. Reprodukcí vnímaného dále ovlivňuje vyjadřovací schopnost svědka.

Na kvalitě sestaveného portréту se významně podílí přesnost poskytnutého popisu pachatele. Schopnosti svědka ovlivňuje jeho vzdělání, povolání, fyzický a zejména psychický stav. Dále jsou schopnosti svědka ovlivněny podmínkami pozorování, denní dobou a ročním obdobím.

V praxi je potřeba dodržovat určité zásady tak, aby byl sestavený portrét co nejkvalitnější. Pokud je více svědků, kteří přišli do kontaktu s pachatelem, je potřeba zajistit, aby společně nediskutovali o popisu pachatele. V případě, že je svědek z prožití události v šoku, je vhodné s provedením výslechu vyčkat.

## **2.3 Metody vytváření identikitu**

Tato část textu je založena na zdrojích [29, 23, 20]. Po zajištění kvalitního popisu pachatele můžeme přistoupit k sestavování portrétu. Pro tento účel existuje několik metod, které mohou být použity k vytvoření co nejpřesnějšího obrazu pachatele. Mezi tyto metody patří grafická metoda, plastická metoda, fotorobot, fotomontáž, metoda skládaného portrétu a výpočetní techniky.

### **Grafická metoda**

Jedná se o vytvoření podobizny pachatele grafikem na základě slovního popisu svědka. Nevýhodou zde mohou být schopnosti grafika vyhotovit dostatečně kvalitní podobiznu hledané osoby .

### **Plastická metoda**

Na základě výpovědi svědka výtvarník (sochař) vymodeluje hlavu pachatele. Tato metoda je poměrně náročná, jak na schopnosti svědka, tak na schopnosti výtvarníka. V kriminalistické praxi se téměř nevyužívá.

### **Fotorobot**

U metody fotorobot se po úvodním popisu morfologických znaků obličeje svědkovi předloží tzv. album typů, ve kterém si vybírá obecné obličeje odpovídající vzhledu hledané osoby. Následně je využito kompoziční album, ve kterém jsou fotografie obličejů rozstříhány do

šesti pruhů, které obsahují jednotlivé části obličeje. Znaky, se kterými svědek nesouhlasil na fotografii z alba typů, postupně nahrazuje jinými částmi obličeje z kompozičního alba. Tento proces je třeba několikrát opakovat, dokud svědek s portrétem nesouhlasí.

### **Fotomontáž**

Jedná se o metodu, která vychází z metody fotorobot a je též označována jako mozaikový portrét. Jednotlivé prvky z kompozičního alba jsou zachyceny na šesti diapásech, přičemž každý z nich představuje jednu konkrétní část obličeje. Tyto prvky se následně promítají pomocí šesti samostatných projektorů. U metody fotomontáž je navíc možné upravovat rozdílné velikosti jednotlivých částí obličeje, čímž lze dosáhnout realističtějšího výsledku.

### **Metoda skládaného portréту (Identikit)**

Portrét pachatele je skládán z průsvitných fólií, na kterých jsou vyobrazeny jednotlivé části obličeje např. vlasy, čelo, nos, oči, brada, rty, obočí a vousy. Svědek po provedeném výslechu vybírá ze vzorkovnice jednotlivé části obličeje, které policejní specialista skládá do výsledného portréту. Portrét se případně dokresluje částmi, které neodpovídají vzhledu. Jednotlivé fólie jsou označeny číselným kódem.

### **Výpočetní technika**

V současnosti se portréty osob často sestavují pomocí výpočetní techniky. Tato metoda využívá specializované programy, které umožňují rychlé vytvoření portréту na základě popisu svědka. Tyto aplikace obvykle obsahují databáze obličejových rysů, které lze upravovat a kombinovat pro co nejpřesnější výsledek. Výhodou je možnost přizpůsobení a úpravy portréту v průběhu procesu. Konkrétní aplikace používané v praxi jsou uvedeny v sekci 2.4.

## **2.4 Aplikace pro vytváření identikitu**

Existují různé aplikace, které pomáhají sestavit podobu hledané osoby, a to jak klasické databázové systémy, tak moderní nástroje využívající umělou inteligenci.

### **2.4.1 PORIDOS**

PORIDOS (PORtrétní Identifikace OSob) [23] je software, jehož prostřednictvím se z fotografické databáze obličejových partií sestavuje portrét hledané osoby. Využívají se dílčí části obličeje (např. oči, nos, ústa) a doplňky (např. brýle, vousy). Jednotlivé obličejové komponenty lze různě kombinovat a upravovat v editoru. Výsledný obrázek lze následně vytisknout. Jedná se o program vyvinutý v roce 1992 na Kriminologickém ústavu v Praze. Poridos je provozován ve dvou základních verzích, a to starší černobílá a novější barevná verze s označením wPORIDOS, která obsahuje i asijské typy. Předností systému je poměrně vysoká kvalita sestavených portréů a přehledné menu umožňující rychlý výběr obličejových částí v katalogu. Obličejový katalog lze dále upravovat a rozšiřovat.

Portrét lze sestavovat v režimu kompozice obrazu, kde se za pomoci obrazového menu vybere vhodný tvar obličeje. Další části obličeje lze vybrat obdobně. V programu je možné měnit polohu jednotlivých prvků obličeje, včetně jejich jasu a kontrastu. Korekci lze též provádět v automatickém režimu. Další úpravy se provádí v režimu editace obrazu. Za pomoci funkce retuš je možné vyhladit vzájemné přechody jednotlivých prvků obličeje. Posuv a deformace jsou funkce, s jejichž pomocí lze plynule měnit proporce libovolných částí obrazu. Při použití kombinace kreslení ploch, kreslení čar a retuše je možné dokreslovat vlasy, vousy nebo jiné znaky.

## 2.4.2 FACES

Software FACES [10], vyvinutý společností IQ Biometrix, byl představen v roce 1998. Je to pokročilý nástroj pro tvorbu identikitů podezřelých, který využívají policejní složky mnoha zahraničních států. Jeho databáze obsahuje více než 4 400 prvků obličeje, včetně tříbarevných variant vlasů s možností zrcadlového otočení účesu, přidávání jizev, mateřských znamének, piercingů a tetování. Uživatelé mohou také využít funkce pro úpravu věku a přidávání pokrývek hlavy. Funkce přiblížení a mřížky pro přesné umístění prvků obličeje zajišťují vysokou přesnost při tvorbě identikitů.

FACES byl využíván v pořadu *America's Most Wanted* na stanici Fox, kde přispěl k dopadení několika hledaných osob.

Uživatelské rozhraní aplikace FACES je zobrazeno na obrázku 2.1.



Obrázek 2.1: Aplikace FACES [19].

## 2.4.3 DeepFaceDrawing

DeepFaceDrawing [8] je systém, který umožňuje generování realistických portrétů na základě ručně kreslených skic. Uživatel zadává pouze jednoduchý náčrt obličeje obsahující základní linie očí, nosu, úst a tvar hlavy. Systém následně automaticky vytváří realistický obrázek lidské tváře, který odpovídá stylu i rysům předložené kresby. Výsledný portrét je plně barevný a vykazuje vysokou vizuální kvalitu, a to i v případě velmi hrubého nebo nepřesného vstupního náčrtu.

Program využívá technologii hlubokého učení a je postaven na kaskádě neuronových sítí. Celý proces generování probíhá ve třech fázích. Nejprve je náčrt převeden do tzv. strukturovaného latentního prostoru, ve kterém jsou identifikovány jednotlivé rysy obličeje. Následuje převod těchto rysů do fotorealistické podoby pomocí speciálně navrženého generátoru založeném na GAN (generativní soupeřící síť, viz sekce 3.3). V závěrečné fázi se obraz vylepšuje do realistické podoby.

Výhodou systému DeepFaceDrawing je schopnost pracovat i s velmi nedetailními nebo nepřesnými kresbami, čímž se výrazně zjednodušuje uživatelské ovládání. Uživatel se nemusí soustředit na realistické kreslení, systém sám automaticky doplňuje detaily a vytváří odpovídající rysy obličeje.

Architektura systému je navržena modulárně a umožňuje dále upravovat či kombinovat jednotlivé fáze generování. Uživatel může například interaktivně upravovat náčrt a okamžitě sledovat změny v generovaném portrétu.

## 2.5 Kognitivní funkce ovlivňující tvorbu identikitu

Kognitivní funkce, jako je vnímání a paměť, zásadně ovlivňují přesnost tvorby identikitu. Schopnost pamatovat si a přesně popsat detaily obličeje může být ovlivněna různými psychologickými procesy, což vede k chybám v identifikaci.

### 2.5.1 Paměť a vnímání obličeje

Vzpomínky, které si uchováváme, nejsou vždy přesné a mohou být zkreslené. To platí obzvláště v případě stresových situací, jako je například trestný čin, kdy dochází k ovlivnění vnímání. Popisující osoba si pak může situaci vybavit jinak, než jak ve skutečnosti probíhala. V praxi se také stává, že pokud více svědků před výpovědí společně rozebírá danou situaci, může to negativně ovlivnit jejich subjektivní vnímání a vést k nepřesnému popisu.

Lidé si lépe pamatují hlavní rysy nebo výrazné znaky (například jizvy nebo mateřská znaménka), zatímco detaily, jako tvar nosu nebo očí, bývají zapomenuty nebo doplňovány podle předchozích zkušeností či stereotypů [23].

Další problematickou skutečností je tzv. *cross-race efekt*, který popisuje, že svědci mají tendenci hůře rozpoznávat pachatele jiné etnické skupiny, než je jejich vlastní. Tento jev může vést k chybným identifikacím [16].

### 2.5.2 Verbální popis a jeho limity

Převod vizuální informace do slovního popisu je komplikovaný proces. Mnoho lidí nemá bohatou slovní zásobu k popisu obličeje, a tím vytváří nepřesné popisy (např. „měl zlé oči“, „byl nesympatický“), které jsou pro tvorbu portréту problematické. I ten nejlepší úřední nebo laický popis obličeje nemůže ve slovní formě poskytnout přesný obraz konkrétní osoby, protože různí lidé si pod stejným slovním vyjádřením mohou nezávisle na sobě představovat odlišný vzhled popisovaných částí obličeje [20].

### 2.5.3 Rozpoznávání tváří

Studie *Parts and wholes in face recognition* [30] zkoumá, zda lidé rozpoznávají tváře spíše jako celek, nebo prostřednictvím jednotlivých částí. Autoři provedli tři experimenty, aby tuto otázku objasnili:

#### Experiment 1

Účastníci se učili rozpoznávat dvě sady tváří, a to běžné a zamíchané, kde byly části tváře přeuspořádány. Po fázi učení následoval test, ve kterém měli identifikovat konkrétní části obličeje (oči, nos, ústa) buď izolovaně, nebo v kontextu celé tváře. Výsledky ukázaly, že u běžných tváří byli účastníci přesnější při identifikaci částí obličeje v kontextu celé tváře než při jejich izolovaném zobrazení. U zamíchaných tváří tento rozdíl nebyl tak výrazný,

což naznačuje, že normální tváře jsou zpracovávány holisticky (jako celek), zatímco u zamíchaných tváří je zpracování více zaměřeno na jednotlivé části.

### **Experiment 2**

Druhý experiment zkoumal, jak změna orientace (zrcadlení) tváře ovlivňuje její rozpoznávání. Účastníci měli za úkol zapamatovat si přirozeně orientované a převrácené tváře a následně identifikovat části obličeje buď izolovaně, nebo v celém kontextu obličeje, stejně jako v experimentu 1. U převrácených obličejů nebyl mezi úspěšností rozpoznávání izolovaných komponentů a těch, které byly prezentovány v kontextu celé tváře, velký rozdíl. Naproti tomu u obličejů s přirozenou orientací bylo rozpoznávání izolovaných částí méně úspěšné. To naznačuje, že holistické vnímání se uplatňuje pouze u obličejů v přirozené poloze.

### **Experiment 3**

Třetí experiment porovnával rozpoznávání tváří s rozpoznáváním jiných objektů, například domů. Účastníci byli testováni na identifikaci částí a celků u obou typů objektů. Zatímco u tváří byla identifikace celků přesnější než identifikace jednotlivých částí, u domů nebyl takový rozdíl zaznamenán. To naznačuje, že holistické zpracování je specifické pro tváře.

Celkově studie podporuje myšlenku, že lidský mozek zpracovává tváře jako celky, což je odlišuje od jiných objektů, kde je zpracování více zaměřeno na jednotlivé části.

## Kapitola 3

# Generování snímků obličeje pomocí neuronových sítí

Generování syntetických obličejů pomocí neuronových sítí dnes představuje významnou oblast umělé inteligence, která nachází využití například v zábavním průmyslu nebo marketingu. Při vytváření snímku obličeje se využívají generativní modely GAN (generativní soupeřící sítě), VAE (variační autoenkodér), autoregresní a difúzní. Hlavním cílem těchto modelů je vygenerovat data, která mohou být vytvořena na základě určitého vstupu, nebo zcela náhodně [1].

### 3.1 Úvod do neuronových sítí

Neuronová síť je model strojového učení, který se učí plnit úkoly rozpoznáváním vzorců v trénovacích datech. Trénovací data jsou obvykle předem označena ručně. Například v systému, jehož úkolem je rozpoznávat objekty na obrázcích, můžeme trénovat neuronovou síť na sadě obrázků s označením jednotlivých částí.

Neuronová síť je inspirována biologickými procesy v lidském mozku. Je tvořena z tisíců, nebo dokonce miliónů jednoduchých procesních uzlů, nazývané **neurony**, které přijímají vstupní data a následně je zpracovávají. Neurony jsou typicky organizovány do vrstev a fungují na principu dopředné neuronové sítě (angl. *feed forward*), což znamená, že data prochází neuronovou sítí pouze od vstupu k výstupu. Jednotlivé neurony jsou propojeny s neurony v nižší vrstvě, od kterých data přijímají, a zároveň s neurony ve vyšší vrstvě, kam data odesílají. Ke každému propojení mezi dvěma neurony je přidělena hodnota tzv. váha, která udává, jak velký vliv má výstup jednoho neuronu na vstup druhého [14].

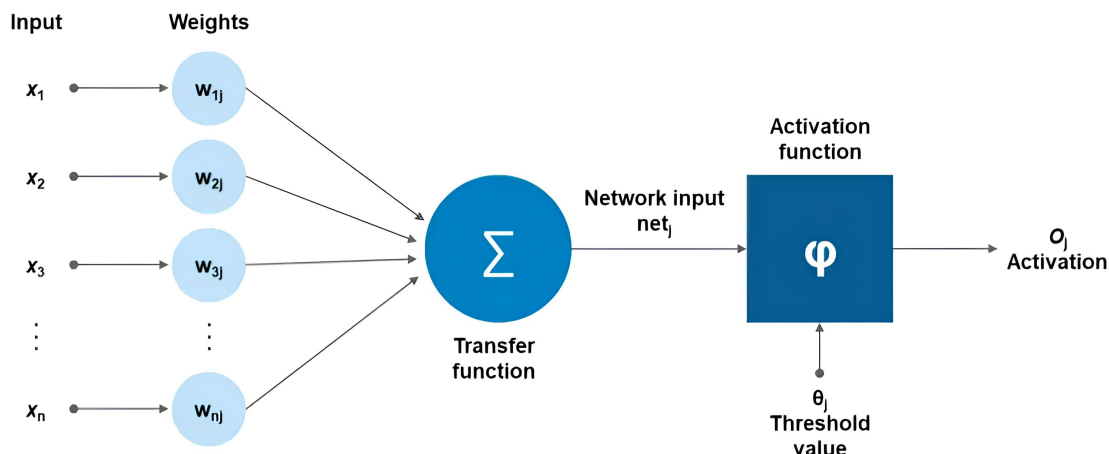
### 3.2 Model neuronu

Umělý neuron je základní stavební jednotkou neuronových sítí, který napodobuje principy biologického neuronu. Biologické neurony jsou tvořeny ze tří základních částí [6, 32]:

- **Dendrity** – Krátké a silně větvené výběžky, které slouží neuronům k přijímání nervových signálů od ostatních neuronů.
- **Tělo neuronu** – Obsahuje jádro a probíhá zde tvorba proteinů. V těle neuronu se zpracovávají signály přijaté dendrity.

- **Axon** – Výběžek neuronu, který slouží k přenosu nervových impulzů z těla neuronu směrem k jiným neuronům.

Podobně jako biologický neuron, umělý neuron přijímá vstupy, zpracovává je a na základě těchto výpočtů generuje výstup. Struktura umělého neuronu je zobrazena na obrázku 3.1.



Obrázek 3.1: Umělý neuron [6].

Na obrázku 3.1 je vyobrazen umělý neuron, který obsahuje vstupy  $x_1, x_2, x_3, \dots, x_n$  reprezentující biologické dendrity. Každý vstup  $x_i$ , kde  $i = 1, 2, \dots, n$ , má přidělenou váhu  $w_i$ , která upravuje danou vstupní hodnotu a tím udává důležitost vstupu. Tyto váhy reprezentují synaptické spoje v biologickém neuronu. Po zpracování všech vážených vstupů, neuron provádí jejich součet pomocí přenosové funkce. K součtu vážených vstupů se přičte konstanta  $b$  tzv. *bias*. Výsledkem tohoto výpočtu je tzv. *síťový vstup*, který slouží jako základ pro další krok. Aby neuron rozhodl, zda aktivovat svůj výstup, je aplikována aktivační funkce. Tato funkce vyhodnotí, zda hodnota síťového vstupu překračuje předem stanovenou prahovou hodnotu. Pokud hodnota překračuje práh, neuron se aktivuje, což znamená, že vyšle svůj výstup dál do dalších neuronů v síti. Tento stav se označuje jako aktivovaný neuron. Naopak, pokud je hodnota síťového vstupu menší, než je prahová hodnota, neuron se neaktivuje a výstup zůstane nulový. Chování umělého neuronu je možno matematicky zapsat vzorcem 3.1 [6, 32].

$$y = f\left(b + \sum_{i=1}^n x_i w_i\right) \quad (3.1)$$

### 3.2.1 Aktivační funkce

Existuje několik typů aktivačních funkcí, z nichž každá má specifické vlastnosti a oblasti vhodného použití.

#### Skoková funkce

Skoková aktivační funkce je jednou z nejjednodušších aktivačních funkcí používaných v neuronových sítích. Pokud je vstupní hodnota vyšší než určitý prahový bod, výstup je 1, jinak je výstup 0. Funkce je definovaná předpisem 3.2.

$$F(x) = \begin{cases} 0, & \text{pro } x < \text{prahová hodnota} \\ 1, & \text{pro } x \geq \text{prahová hodnota} \end{cases} \quad (3.2)$$

Skoková funkce není diferencovatelná (tj. nemá derivaci v určitém bodě), což způsobuje dva hlavní problémy, které brání jejímu použití v hlubokých neuronových sítích. Prvním problémem je, že algoritmus zpětného šíření chyby (angl. *backpropagation*), který upravuje váhy mezi neurony na základě chyb, vyžaduje diferencovatelné aktivační funkce. Bez derivace není možné efektivně vypočítat gradient (určuje se podle derivace aktivační funkce na vstupu neuronu) a upravovat váhy. Druhým problémem jsou skokové změny ve výstupech, které způsobují, že malé změny ve vstupu mohou vést k náhlým skokům ve výstupech neuronů.

### Lineární funkce

Lineární funkce je další z jednoduchých aktivačních funkcí, která se používá v neuronových sítích. V tomto případě je výstup neuronu přímo úměrný vstupu. Funkce je daná rovnicí 3.3.

$$F(x) = x \quad (3.3)$$

Lineární funkce se využívají zejména u regresních úkolů, jejichž prostřednictvím dochází k predikci spojitéch hodnot.

### Sigmoidní funkce

U sigmoidní funkce se výstup pohybuje v intervalu mezi 0 a 1, což ji činí vhodnou pro úkoly, kde je potřeba výstup vyjádřit jako pravděpodobnost. Sigmoidní funkce je definována předpisem 3.4 [3].

$$F(x) = \frac{1}{1 + e^{-x}} \quad (3.4)$$

Pokud výstup sigmoidní funkce nabývá velmi vysokých nebo nízkých hodnot, jeho derivace je velmi malá. To znamená, že při zpětném šíření chyby je gradient velmi slabý, což výrazně zpomaluje proces úpravy vah neuronů.

### Hyperbolický tangens

Hyperbolický tangens je nelineární aktivační funkce, která se podobá sigmoidní funkci, ale její výstupy jsou v intervalu mezi -1 a 1 [3]. Tato funkce je definována funkcí 3.5.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.5)$$

Díky tomu, že výstupy mohou být jak kladné, tak záporné, hyperbolický tangens pomáhá při modelování signálů, které mají symetrické rozdělení kolem nuly.

### ReLU

ReLU (Rectified Linear Unit) [3] je jedna z nejpoužívanějších aktivačních funkcí používaných v neuronových sítích. ReLU je definována funkcí 3.6.

$$F(x) = \max(0, x) \quad (3.6)$$

Funkce pro všechny kladné hodnoty vstupu vrací jeho hodnotu, zatímco pro záporné hodnoty vrací nulu. Tato jednoduchá funkce umožňuje rychlé a efektivní trénování neuronových sítí, protože její derivace je buď 0 nebo 1.

Existují různě upravené verze aktivační funkce ReLU, například Leaky ReLU a Exponential Linear Unit, které brání rozpadu gradientu. Tento problém vzniká, pokud gradient během zpětného šíření klesá exponenciálně blízko k nule [3].

### 3.2.2 Principy učení neuronové sítě

Během tréninku se mění hodnoty vah a prahů v modelu tak, aby se výstupy přibližovaly k požadovaným hodnotám.

**Učení s učitelem** pracuje s označenými daty, což znamená, že ke každému vstupu musí být přiřazen správný výstup, obvykle ručně připravený člověkem. Model se učí rozpoznávat vzory mezi vstupy a výstupy, což mu následně umožňuje předpovídat nové výstupy. Tento proces je přesný, ale časově velmi náročný.

**Učení bez učitele** naopak pracuje s neoznačenými daty. Model se často využívá k identifikaci vzorů v datech a k jejich následnému seskupení na základě podobnosti. Tento přístup je méně náročný na lidskou práci, ale může být obtížnější interpretovat jeho výsledky [5].

## 3.3 Generativní soupeřící sítě

Generativní soupeřící sítě (GAN) [24, 13] představují inovativní přístup v oblasti strojového učení, který se zaměřuje na generování nových dat, zejména fotorealistických obrázků. GAN jsou tvořeny dvěma neuronovými sítěmi: **generátorem** a **diskriminátorem**. Tyto dvě části sítě mají za úkol spolupracovat v prostředí, kde vzájemně soupeří s cílem zlepšit své schopnosti, což vede k vytvoření vysoce kvalitních dat.

### Generátor

Generátor je zodpovědný za tvorbu nových dat, například obrázků nebo textů, která by měla co nejvíce připomínat skutečná data. Generátor začíná tím, že vygeneruje náhodný šum, který následně prochází neuronovými vrstvami, kde se postupně transformuje do výstupu, který má být co nejvíce realistický. Tento výstup je odeslán na diskriminátor, který rozhodne, zda se data podobají reálným. Pokud diskriminátor rozhodne, že data nejsou validní, generátor upraví své váhy, aby vylepšil výstup pro další iteraci. Tento proces pokračuje, dokud generátor nezačne produkovat požadovaná data [24].

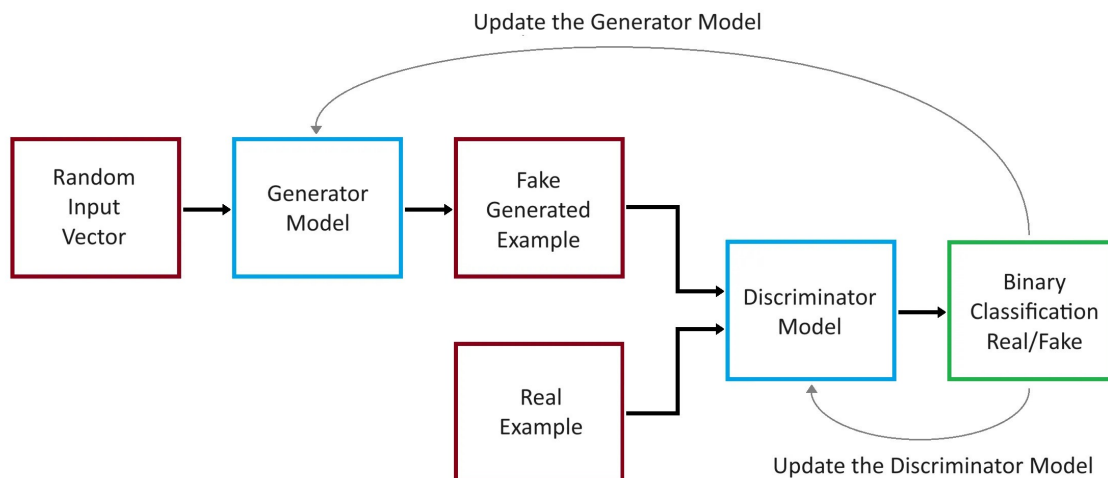
### Diskriminátor

Diskriminátor je naopak síť, která se učí odlišovat mezi skutečnými a generovanými daty. Jeho úkolem je hodnotit nová data, která obdrží od generátoru, a určit, zda vypadají realisticky, nebo byla vytvořena generátorem. Diskriminátor je v podstatě klasifikační síť, která se trénuje na základě označených dat. Pokud diskriminátor špatně označí generovaná data jako skutečná, nebo skutečná data jako falešná, dostane zpětnou vazbu, která vede k jeho vylepšení.

### Trénování GAN

Trénování GAN využívá princip učení s učitelem pro diskriminátor, který je trénován na základě skutečných a generovaných dat, zatímco generátor využívá metodu posilovaného učení, kde je jeho cílem postupně zlepšovat generování dat. V praxi to znamená, že gene-

rátor se pokouší oklamat diskriminátor, zatímco diskriminátor se snaží být stále chytřejší v rozpoznávání těchto podvržených dat. Proces trénování je znázorněn na obrázku 3.2 [24].



Obrázek 3.2: Architektura trénování modelu GAN [24].

### Příklady algoritmů založených na GAN

Od představení základní architektury GAN vznikla řada rozšíření, která reagují na specifické problémy spojené s trénováním nebo cílovým využitím těchto sítí. Jedním z nejpopulárnějších je **DCGAN** (Deep Convolutional GAN), který využívá konvoluční vrstvy k efektivnímu zpracování obrazových dat a jeho architektura se stala základem pro mnoho moderních variant GAN.

**CGAN** (Conditional GAN) zavádí možnost ovlivnit výstup sítě pomocí přídavné informace (např. textového popisu), kterou generátor i diskriminátor obdrží jako vstup, čímž se síť učí generovat cílený obsah odpovídající dané podmínce.

Významnou a v současnosti velmi populární architekturou je **StyleGAN**, která umožňuje generovat vysoce realistické obrázky, včetně realistických obličejů, a nabízí pokročilé řízení vizuálních vlastností výsledného obrázku prostřednictvím kontroly různých faktorů, jako jsou vlasy, věk nebo pohlaví [13].

## 3.4 Variační autoenkodér

Variační autoenkodér (VAE) [2] patří mezi generativní modely strojového učení, který generuje nová data, podobná těm, na kterých byl model trénován. Klíčovou vlastností VAE je schopnost zachytit složité vztahy a struktury v trénovacích datech do tzv. **latentního prostoru** a na základě těchto vztahů generovat nová data. Struktura VAE se skládá ze dvou hlavních částí, a to z enkodéru a dekodéru.

### Enkodér

Enkodér se zaměřuje na převod vstupních dat do jejich latentní reprezentace. Na rozdíl od tradičních autoenkodérů, které generují pevné body v latentním prostoru, VAE vytváří distribuci s parametry, které jsou obvykle rozloženy podle Gaussovy funkce. Pro každý latentní atribut trénovacích dat se kódují dva vektory, a to střední hodnota a směrodatná odchylka,

které tuto distribuci definují. Tímto způsobem si model zachovává originalitu, protože latentní prostor není omezen na pevné body, ale umožňuje generování různých vzorků, které stále zachovávají klíčové vlastnosti trénovacích dat.

### Dekodér

Klíčová funkce pro dekodér je generování dat. Dekodér se chová jako opak enkodéru, to znamená, že bere náhodný vzorek z distribuční reprezentace vytvořené enkodérem a přetváří jej zpět do původního formátu dat. Tento proces umožňuje modelu nejen rekonstruovat data, ale také generovat nové, dosud neexistující vzorky, které zachovávají charakteristiky trénovacích dat.

### Trénování VAE

Trénování VAE modelu spočívá v optimalizaci speciální ztrátové funkce, která se skládá ze dvou částí.

První část je rekonstrukční ztráta, která měří, jak dobře model dokáže obnovit původní data z jejich latentní reprezentace. K tomu se obvykle používá metoda Mean Squared Error (MSE), která penalizuje rozdíly mezi původními daty a jejich rekonstrukcí.

Druhou částí je Kullback-Leiblerova divergence, která zajišťuje, že distribuce latentního prostoru bude co nejvíce podobná Gaussovému rozdělení. Tato složka působí jako kontrolní mechanismus, který pomáhá udržet latentní prostor organizovaný a zabraňuje modelu, aby se příliš přizpůsobil konkrétním trénovacím datům. Optimalizace probíhá prostřednictvím zpětné propagace, kde se váhy enkodéru a dekodéru iterativně upravují tak, aby obě složky ztráty byly co nejnižší [2].

### Příklady algoritmů založených na VAE

VAE se staly základem pro řadu variant, které řeší specifické problémy generování a modelování latentních prostorů. **Beta-VAE** [33] se zaměřuje na oddělení různých faktorů v latentním prostoru. Pomocí nastavení speciálního parametru model vyvažuje schopnost ukládat informace a zajišťuje, že se každý faktor učí nezávisle. To umožňuje, aby model rozpoznával a generoval jednotlivé vzory obsažené v datech.

**CVAE** (Conditional VAE) [2] umožňuje generovat data na základě podmíněných vstupů, jako jsou textové popisy, což usnadňuje cílené generování specifických dat.

**VQ-VAE** (Vector Quantized Variational Autoencoder) [33] je typ autoenkodéru, který místo běžného kontinuálního latentního prostoru používá diskrétní hodnoty. To znamená, že model mapuje informace do konečné množiny vektorů, což zlepšuje kvalitu generovaných obrazů nebo jiných dat a činí je ostřejšími a realističtějšími.

## 3.5 Difúzní modely

Difúzní modely [9] jsou typem generativních modelů, které se ukázaly jako velmi efektivní při generování realistických obrázků a dalších úloh počítačového vidění. Tento model je založen na procesu, který simuluje šíření šumu do datového vzorku a následně se snaží tento šum rekonstruovat zpět do původního formátu. Klíčovou vlastností difúzních modelů je schopnost generovat nové vzorky dat pomocí postupu, který zahrnuje postupnou přeměnu šumu na strukturovaná data. Tento proces je řízen stochastickými procesy, které jsou postupně obráceny, aby vytvořily nový datový vzorek. Stochastický proces je náhodný proces řízený pravděpodobnostními pravidly, kde výsledek každého kroku závisí na náhodě a předchozím stavu.

## Difúzní proces

Jednou z hlavních částí difúzního modelu je proces přidávání šumu k trénovacím datům. Tento šum je postupně přidáván k původním datům v několika krocích, což je modelováno pomocí určitého stochastického procesu. V každém kroku se data mění a stávají se více náhodná, až nakonec dosáhnou stavu, který je téměř identický s náhodným šumem. Tento proces se používá jako základ pro trénování modelu, který se naučí reverzní proces, neboli jak rekonstruovat původní data z tohoto šumu.

V difúzním procesu [9] je stav  $x_t$  v čase  $t$  definován následující rovnicí, která kombinuje původní signál  $x_0$  a přidaný šum  $\epsilon_0$ :

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon_0 \quad (3.7)$$

kde:

- $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$  je parametr, který určuje, jak moc původní signál zůstává v obraze  $x_t$ , kde  $\beta_s$  je míra šumu přidávaného v každém kroku,
- $x_0$  jsou původní data, která se snažíme rekonstruovat,
- $\epsilon_0$  je přidaný šum v prvním kroku procesu.

## Reverzní proces

Klíčovou funkcí difúzního modelu je schopnost obnovit původní data ze šumu pomocí procesu zpětné difúze. Tento proces se trénuje na základě rekonstruování dat od šumu zpět k originálním datovým vzorkům. V praxi model využívá reverzní difúzní proces, který v každém kroku předpovídá a odstraňuje šum přítomný v datech, čímž postupně rekonstruuje původní vzorek. Tento proces umožňuje generování nových dat tím, že krok za krokem transformuje šum do strukturovaných dat, které odpovídají vzorcům v trénovacích datech.

## Trénování difúzních modelů

Trénování difúzních modelů spočívá v optimalizaci procesu, rekonstrukce původních dat z šumu. Tento proces zahrnuje dvě hlavní fáze.

V první fázi (přední difúze) se do dat přidává šum až do úplného zničení původních informací. Ve fázi zpětné difúze model postupně odstraňuje tento šum a učí se rekonstruovat originální data. Během trénování model optimalizuje své parametry na základě rozdílů mezi skutečnými a rekonstruovanými daty.

Cílem je minimalizovat chyby mezi šumem přidaným k původním datům a tím, co model na základě tohoto šumu vygeneruje. Pomocí stochastické gradientní optimalizace a zpětné propagace model postupně zlepšuje svou schopnost generovat realistická data, která odpovídají trénovacím vzorkům [9].

## Příklady algoritmů založených na difúzních modelech

Mezi hlavní směry vývoje difúzních modelů patří **DDPMs** (Denoising Diffusion Probabilistic Models), které obnovují data krok za krokem tak, že postupně odstraňují přidaný šum pomocí pravděpodobnostního postupu.

Jiným přístupem jsou **score-based generative models**, které data rekonstruuji na základě toho, jak pravděpodobné jsou různé tvary dat, aby odpovídala reálnému rozdělení.

Efektivní variantou je **LDM** (Latent Diffusion Model), který generuje data v latentním prostoru, což výrazně snižuje výpočetní náročnost. Tuto architekturu využívá například model **Stable Diffusion** [26, 9].

## 3.6 Autoregresivní model

Autoregresivní modely jsou klíčovou technikou v analýze a predikci časových řad, které využívají hodnoty z předchozích časových okamžiků k předpovědi budoucích hodnot. Autoregresivní modely jsou vysoce efektivní, zejména v případech, kdy jsou v datech jasně patrné vzory a závislosti mezi časovými kroky.

Autoregresivní modely generují data sekvenčně tak, že každý bod v sekvenci je predikován jako pravděpodobnostní distribuce závislá na předchozích bodech (vzorec 3.8).

$$p(x) = \prod_{i=1}^n p(x_i | x_1, x_2, \dots, x_{i-1}) \quad (3.8)$$

Kde:

- $x_i$  je aktuální hodnota, která závisí na předchozích hodnotách  $x_1, x_2, \dots, x_{i-1}$ ,
- $p(x_i | x_1, x_2, \dots, x_{i-1})$  je distribuce, podle které model předpovídá hodnotu  $x_i$ , přičemž tato distribuce je podmíněná na všech předchozích hodnotách sekvence.

Proces generování začíná náhodně inicializovaným stavem, nebo je stanoven na základě specifických požadavků. Dále model iterativně předpovídá hodnoty na základě předchozích predikcí a postupně buduje kompletní sekvenci. Každá nová hodnota  $x_i$  je generována z distribuce  $p(x_i | x_{<i})$ , která byla naučena během trénování.

### Trénování autoregresivního modelu

Trénování autoregresivního modelu je většinou rozděleno do dvou částí. První částí je předtrénování bez dozoru. U předtréninku se snaží model maximalizovat logaritmickou pravděpodobnost následujících dat. Například u generování obrázků se model snaží co nejlépe predikovat následující vizuální objekt. Po předtrénování na neoznačených datech se model učí pomocí označených dat. Vstupy (např. obrázky) jsou předávány přes již předtrénovaný model do lineární vrstvy, která predikuje cílové hodnoty, které jsou označené v trénovacích datech. Během této fáze se optimalizuje nejen hlavní cíl, ale i pomocný cíl z předtrénování [7].

### Příklady algoritmů založených na autoregresivních modelech

Mezi varianty autoregresivních modelů patří **VAR** (Vector Autoregressive Models), které umožňují modelování vícerozměrných časových řad, kde každá proměnná závisí nejen na svých minulých hodnotách, ale i na minulých hodnotách ostatních proměnných.

Dalšími známými rozšířeními jsou **ARMA** a **ARIMA**, které kombinují autoregresi s klouzavými průměry. Klouzavý průměr v těchto modelech modeluje chybové členy jako kombinaci předchozích hodnot, což vyhlazuje data a zlepšuje predikci.

Pro modelování proměnlivého rozptylu se používají modely **ARCH** a **GARCH**, které predikují volatilitu v časových řadách s měnícím se rozptylem [21].

## 3.7 Text-to-image modely

Text-to-image modely umožňují generovat obrázky na základě textového popisu. Základem těchto modelů je textový kodér, který analyzuje přirozený jazyk a převádí ho do vektorové reprezentace. Tyto proměnné jsou základem pro generativní model. Nejvyužívanější přístup pro generování obrázků na základě textu je v poslední době difúzní model [34].

### 3.7.1 Stable Diffusion

Stable Diffusion je pokročilý open-source text-to-image model, který byl vyvinut společností Stability AI. Je založen na latentním difúzním modelu, který využívá difúzní proces pro vytvoření latentních proměnných z textového vstupu. Latentní prostor ve Stable Diffusion je 48x menší než samotný obrazový prostor, tudíž difúzní proces v tomto prostoru výrazně zrychluje generování a klade nižší nároky na výpočetní výkon. Stable Diffusion se skládá ze tří částí: textový kodér, U-Net, VAE.

#### Textový kodér

Textový kodér převádí text na vektorové proměnné, které jsou dále využity k řízení procesu generování. Stable Diffusion využívá již natrénovaný jazykový model CLIP.

#### U-Net

U-Net je neuronová síť, která se používá k rekonstrukci latentních proměnných z jejich zašuměné podoby. Na vstupu přijímá zašuměné latentní proměnné a výstup textového kodéru. Během svého procesu postupně odstraňuje šum, čímž formuje latentní proměnné tak, aby odpovídaly vstupnímu textu. Tento proces se několikrát opakuje, dokud šum zcela nezmizí. V novějších verzích Stable Diffusionu 3.0 a 3.5 byl U-Net nahrazen architekturou difúzního transformátoru, který vylepšuje schopnost porozumění textu.

#### VAE

Při generování obrazů v Stable Diffusion, VAE dekoduje latentní proměnné získané z U-Net zpět do plnohodnotného obrazu. Dekódování pomocí VAE zajišťuje, že výsledné obrazy jsou detailní a díky tomu lze provádět difúzi v menším latentním prostoru [34, 35, 26].

### 3.7.2 Dall-E 2

Dall-E je proprietární text-to-image model vyvinutý společností OpenAI. Jeho první verze byla založena na autoregresivním modelu, ale ve druhé verzi z důvodu generace kvalitnějších snímků byl převeden na difúzní model. Oproti Stable Diffusion, funguje na principu přímé generace v obrazovém prostoru, což vede k vyšší kvalitě výstupů, ale zároveň klade vyšší nároky na výpočetní výkon. Architektura Dall-E 2 je složena z následujících prvků:

#### Textový kodér

Textový kodér v Dall-E 2 využívá model CLIP, který převádí vstupní text na vektorovou reprezentaci.

#### Textový prior

Prior slouží k transformaci textového vektoru z CLIP na odpovídající obrazový vektor. Může být naučen buď pomocí autoregresivní metody, nebo difúzního modelu, přičemž difúzní prior poskytuje lepší výsledky. Bez tohoto prioru by výkon modelu výrazně klesl.

#### Difúzní dekodér

Difúzní dekodér Dall-E 2 na základě obrazového vektoru generuje obrázek. Tento proces funguje na základě difúzního modelu, který umožňuje postupné odstraňování šumu z generovaných obrazů [34, 22].

### 3.7.3 Imagen

Imagen [27] je text-to-image model vyvinutý společností Google, který se zaměřuje na fotorealistické generování obrazů. Na rozdíl od modelů využívajících latentní prostor, Imagen generuje obrazy přímo v pixelovém prostoru pomocí kaskádového difúzního procesu. Architektura Imagenu se skládá ze dvou hlavních komponent: textového kodéru a difúzního generátoru.

#### Textový kodér

Imagen využívá natrénovaný jazykový model T5, který převádí vstupní text do vektorové reprezentace.

#### Difúzní generátor

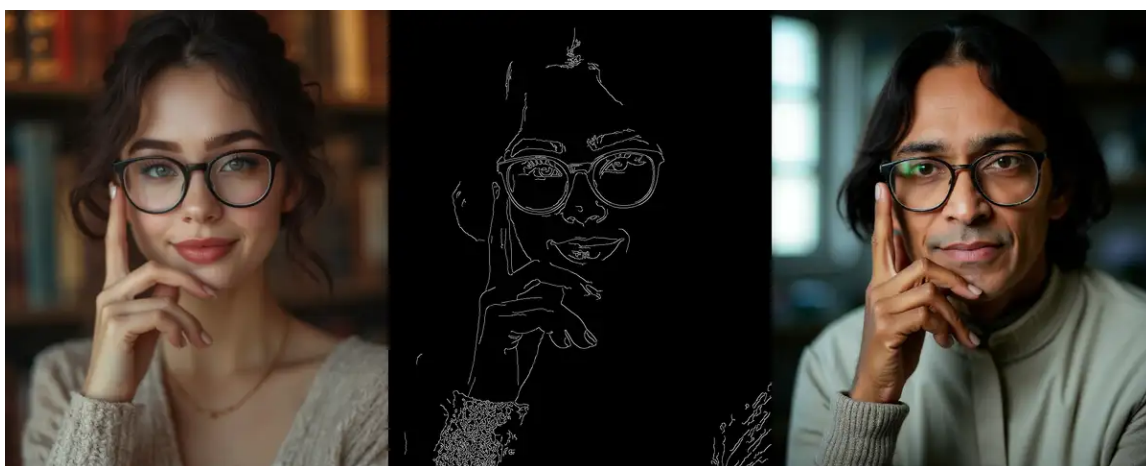
Proces generování v Imagenu probíhá ve třech krocích. Nejprve základní difúzní model vytvoří hrubý obraz s rozlišením  $64 \times 64$ . Tento obraz je následně vylepšován pomocí dvou super-rezolučních difúzních modelů, které výstup postupně zvětšují na  $256 \times 256$  a finálně na  $1024 \times 1024$  pixelů. Tato kaskádová architektura umožňuje produkci vysoce detailních a konzistentních obrazů.

## 3.8 ControlNet

ControlNet [35] je technologie, která dokáže přidávat různé podmínky do velkých předtrénovaných modelů. Na vstup modelu ControlNetu obvykle přichází nějaký obrázek, který definuje podmínky, podle kterých model generuje výstupy. Těmito podmínkami mohou být například obrysy objektů, hloubka scény, pozice těla apod.

Při procesu trénování ControlNet aplikuje speciální architekturu nulové konvoluce. Při této konvoluci jsou v neuronové síti komponenty jako váha a bias nastaveny na nulu, což znamená, že v počátečních fázích trénování neprovádějí žádné změny v datech, a tím se zabraňuje generování nežádoucího šumu.

Na obrázku 3.3 je zobrazen příklad použití obrysu jako vstupu pro generování obrázků pomocí ControlNet Canny. Ve středu obrázku je obrys, který slouží jako vstup. Vlevo a vpravo od něj jsou obrázky vygenerované na základě tohoto obrysu.



Obrázek 3.3: Generování pomocí ControlNet Canny. [28].

### 3.9 Image inpainting

Image inpainting [12], neboli doplňování obrazu, je technika rekonstrukce vizuálních dat, jejímž cílem je nahradit chybějící, poškozené nebo nežádoucí části obrazu takovým způsobem, aby výsledek působil přirozeně a byl konzistentní se zbytkem scény. Tato metoda nachází uplatnění v řadě oblastí, jako je digitální restaurování historických obrazových materiálů, odstraňování objektů z fotografií, oprava starých nebo poškozených snímků. Hlavním cílem inpaintingu je zachování konzistence barev, textur a tvarů.

Přístupy k image inpaintingu lze rozdělit do dvou hlavních kategorií:

#### Sekvenční metody

Sekvenční metody se dělí na **záplatové (patch-based)** a **difúzní** přístupy. Záplatové metody fungují tak, že hledají podobné části obrazu v nepoškozených oblastech a tyto části kopírují do chybějících míst.

Difúzní metody naopak šíří obrazový obsah z okrajů poškozené oblasti směrem do jejího středu.

Tyto přístupy jsou účinné u jednoduchých obrazů, ale u složitějších scén, které obsahují mnoho textur, objektů, nebo kde objekt pokrývá velkou část obrazu, mohou být méně spolehlivé.

#### CNN metody

Metody založené na konvolučních neuronových sítích jsou novější a využívají hluboké učení k dosažení kvalitnějších výsledků. Tyto metody, jako jsou *encoder-decoder* architektury nebo přístupy založené na GAN, dokážou lépe zachytit globální struktury a detaily, které sekvenční metody obtížně rekonstruují. Přístupy založené na CNN často zahrnují trénink na velkém množství dat, což jim umožňuje přesně obnovit chybějící části obrazů, včetně textur i složitějších struktur [12].

## Kapitola 4

# Návrh řešení

Tato kapitola se zaměřuje na návrh, architekturu a technologické řešení aplikace pro generování identikitu na základě textového popisu v českém jazyce. Postupně popisuje specifikaci požadavků, zvolené technologie, použité knihovny a návrh uživatelského rozhraní. Dále je věnována pozornost jednotlivým fázím generování obličeje, včetně důvodů pro použití ControlNetu a masek obličeje.

### 4.1 Specifikace požadavků

Cílem aplikace je umožnit uživateli vytváření snímku obličeje konkrétní osoby na základě vstupního textu v českém jazyce.

Uživatel by měl být schopen zadat textový popis obličeje osoby pomocí přirozeného jazyka. Popis by měl obsahovat klíčové informace o vzhledu dané osoby. Tento textový vstup bude následně zpracován textovým kódérem, který identifikuje jednotlivé prvky popisu, jako jsou pohlaví, barva a délka vlasů, typ očí, nosu, úst, přítomnost vousů, brýlí a dalších charakteristických znaků.

Výstupem bude obličej zobrazený přímo v aplikaci, který bude odpovídat zadanému popisu uživatelem. V případě, že se některé prvky obličeje nebudou podobat popisované osobě, nebo nebude možné prvek vygenerovat, aplikace nabídne možnost její úpravy. Uživatel si tak bude moci vygenerovaný identikit dále přizpůsobit dle vlastního uvážení, například změnit barvu očí, vlasů nebo přidat další prvky (např. vousy).

Aplikace bude dále umožňovat export výsledného obrázku do počítače.

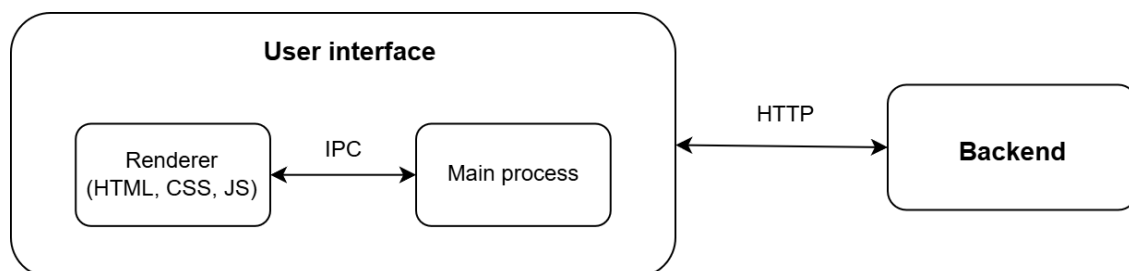
### 4.2 Výběr technologií

Pro realizaci aplikace byla zvolena kombinace programovacího jazyka Python na straně backendu a vývojového prostředí Electron na straně frontendu.

Python byl určen již v zadání bakalářské práce jako hlavní jazyk pro implementaci aplikační logiky. Python je v současnosti jedním z nejpoužívanějších jazyků pro práci s neuronovými sítěmi a díky tomu disponuje širokou nabídkou knihoven pro generativní modely.

Na frontendovou část aplikace byl zvolen framework Electron, který umožňuje vytvářet desktopové aplikace s využitím webových technologií jako je HTML, CSS a JavaScript. Hlavní výhodou Electronu je, že z jednoho kódu lze vytvořit aplikaci, která bude fungovat na různých operačních systémech.

Celková architektura aplikace (obrázek 4.1) je koncipována jako klient-server model, kde frontend aplikace představuje klienta komunikujícího s Python serverem pomocí HTTP požadavků. Uživatelské rozhraní běží lokálně na zařízení uživatele, přičemž backend je spuštěn jako server. Uživatel v Electronu zadává vstupní data (např. textový popis osoby), která jsou následně odeslána na server, kde dochází ke zpracování a vytvoření odpovědi, která je vrácena zpět do uživatelského rozhraní. Takové rozdělení umožňuje oddělit uživatelské rozhraní od výpočetně náročných operací, což je u generování obrázků pomocí neuronové sítě žádoucí.



Obrázek 4.1: Architektura aplikace.

### 4.3 Použité knihovny

Při vývoji aplikace pro generování identikitu bylo klíčové zvolit takové nástroje, které umožní efektivní práci s neuronovými sítěmi, zpracování obrazových dat a překlad textu. Vzhledem k tomu, že byl zvolen programovací jazyk Python, byly použity následující knihovny:

#### Diffusers

Diffusers<sup>1</sup> je knihovna vyvinutá společností Hugging Face, která slouží k práci s difúzními modely. Knihovna nabízí vysokou úroveň abstrakce pro snadnou integraci a experimentování s různými předtrénovanými difúzními modely, jako je například Stable Diffusion a jeho rozšířené verze.

Výhodou knihovny Diffusers je jednoduché a přehledné API, které staví na programovacím jazyce Python a pracuje se standardními knihovnami jako PyTorch. Kromě základních funkcí pro generování obrázků knihovna podporuje i pokročilé techniky, jako je inpainting nebo image-to-image transformace.

Jako nejvhodnější varianta se jeví Stable Diffusion XL (SDXL), která představuje výrazný posun oproti starším verzím modelu. Vyšší verze modelu sice mohou nabízet další vylepšení, avšak jejich praktické využití je v současné době omezené z důvodu velmi nízké dostupnosti modulů ControlNet, které jsou klíčové pro řízení výsledného výstupu.

Základní verze modelu SDXL umožňuje generaci vysoce realistických obličejových rysů s možností stylizace a vysokou mírou detailů. Pro cílené úpravy konkrétních oblastí obrazu je pak vhodné využít nadstavbu SDXL Inpainting, která umožňuje inpainting.

Pro základní verzi modelu SDXL je k dispozici několik modulů ControlNet, které umožňují přesnější definici tvaru obličeje a tím i vyšší míru přizpůsobení výsledného obrazu konkrétního uživatelského záměru.

<sup>1</sup>Dostupné z: <https://huggingface.co/docs/diffusers/index>

## Torch

Torch<sup>2</sup> je základem pro výkonné výpočty ve strojovém učení a je klíčový pro generativní modely jako je Stable Diffusion. Zaměřuje se na práci s tenzory, které slouží k reprezentaci a manipulaci s daty v neuronových sítích. Při generování snímků ve Stable Diffusion umožňuje Torch efektivní numerické operace a využití GPU, což urychluje generování obrazů.

## Deep Translator

Pro dosažení kvalitních výsledků při generování snímků pomocí modelu Stable Diffusion je důležité pracovat s dobře formulovanými vstupními texty, ideálně v anglickém jazyce, na které jsou modely nejlépe trénované. K tomu lze využít knihovnu Deep Translator<sup>3</sup>, která poskytuje jednoduché rozhraní pro překlad textů mezi mnoha světovými jazyky.

## OpenCV, dlib, NumPy

Pro účely detekce obličejových rysů a následné manipulace s jednotlivými částmi obličeje byly zvoleny knihovny OpenCV, dlib a NumPy. Tyto nástroje tvoří osvědčený základ při vývoji aplikací v oblasti počítačového vidění a umožňují efektivní předzpracování vstupních dat pro další generativní kroky.

Knihovna **OpenCV**<sup>4</sup> byla vybrána pro svou širokou funkcionalitu v oblasti zpracování obrazu, včetně základních operací, jako je načítání, transformace a vizualizace obrazových dat. Její použití usnadňuje přípravu vstupního snímku a práci s obrazovými formáty.

**Dlib**<sup>5</sup> slouží v navrženém řešení k detekci klíčových bodů obličeje, jako jsou oči, nos nebo ústa. Tato detekce je zásadní pro možnost cílené úpravy jednotlivých částí obličeje.

**NumPy**<sup>6</sup> pak doplňuje výše uvedené knihovny o podporu efektivních numerických operací s obrazovými daty. Umožňuje práci s obrazem jako s multidimenzionálním polem a slouží k přepočtům souřadnic.

Díky kombinaci těchto tří knihoven je možné snímek obličeje nejen detekovat a rozložit na jednotlivé části, ale také s těmito prvky dále pracovat a připravit je jako vstup pro další generativní zpracování.

## 4.4 Návrh aplikace

Tato část popisuje návrh a koncepci aplikace pro generování identikitu na základě textového popisu v českém jazyce. Aplikace je navržena jako interaktivní nástroj pro tvorbu podoby lidské tváře na základě postupné interpretace slovního zadání a následné grafické úpravy. Cílem aplikace je uživatelsky přívětivým způsobem umožnit tvorbu věrných náčrtků a realistických portrétů osob podle popisu.

Generativní modely, které využívá aplikace, obvykle pracují s anglickými vstupy. Vzhledem k tomu, že uživatelé zadávají popisy v češtině, jsou tyto texty automaticky přeloženy do angličtiny, aby model mohl správně interpretovat požadavky a vygenerovat požadované výsledky.

Na začátku procesu uživatel volí několik parametrů týkajících se počtu vygenerovaných obličejů a úprav. Prvním krokem je volba, kolik variant černobílých skic obličeje bude vygenerováno, kolik úprav přes image inpainting (doplňování či nahrazování části kresby

---

<sup>2</sup>Dostupné z: <https://pytorch.org>

<sup>3</sup>Dostupné z: <https://github.com/nidhaloff/deep-translator>

<sup>4</sup>Dostupné z: <https://opencv.org>

<sup>5</sup>Dostupné z: <https://dlib.net>

<sup>6</sup>Dostupné z: <https://numpy.org/>

pomocí textového popisu) bude k dispozici a kolik realistických obličejů bude vytvořeno v závěrečné fázi. Tato volba umožňuje uživateli přizpůsobit si proces generování podle svých potřeb.

Proces tvorby identikitu je rozdělen do několika navazujících fází, které reflektují přirozený postup od hrubého náčrtu k realistickému ztvárnění. Prvním krokem je zadání jednoduchého popisu jednotlivých rysů obličeje, například tvar obočí, velikost nosu, přítomnost vousů, pohlaví, věk apod. Uživatel vybírá základní masku obličeje, která reprezentuje celkový tvar a proporce hlavy.

Následně je vygenerováno několik variant černobílých skic obličeje. Tyto skici jsou stylizované a slouží jako vizuální návrhy, ze kterých si uživatel může vybrat jednu variantu pro další práci. Vybraný náčrt lze dále upravovat pomocí dvou nástrojů, a to inpaintingu a přímého kreslení do obrazu, které umožňují manuální zásahy, jako je překreslení očí, odstranění brýlí nebo doplnění jizev.

Pro pokročilé úpravy je možné obličej rozložit na jednotlivé komponenty a tyto části samostatně transformovat (měnit jejich polohu a rotaci). Tímto způsobem může uživatel dosáhnout vyšší míry přizpůsobení výsledného výrazu a charakteru postavy.

Závěrečnou fází procesu je převedení finální skici do realistické barevné podoby. Tento krok opět využívá textový popis, například „Mladá žena s modrýma očima, hnědými vlasy a světlou pletí“. Výsledný portrét je možné ještě dále upravit, například doplněním specifických rysů (tetování, mateřská znaménka, piercing apod.) pomocí cíleného inpaintingu nebo kreslení.

Celý proces generování identikitu a interakce mezi jednotlivými komponentami systému jsou znázorněny pomocí sekvenčního diagramu, který je uveden v příloze [B.1](#).

#### 4.4.1 Motivace využití ControlNet a detekce obličejových prvků

Při testování generování obrazu pouze na základě textového popisu, se ukázaly určité limity samotného modelu Stable Diffusion XL. Model často nedokázal přesně vystihnout tvarové rysy obličeje tak, aby odpovídaly konkrétní představě uživatele. Další častou komplikací byla nízká konzistence výstupů. Ve výsledných snímcích se objevovalo celé lidské tělo, nebo naopak příliš přiblížené detaily obličeje, což komplikovalo další zpracování a úpravy. Z tohoto důvodu byl do návrhu začleněn model ControlNet Scribble, který umožňuje generování na základě předem definované skici obličeje. Skica slouží jako přesné vodítko pro zachování tvaru hlavy a rozložení obličejových rysů. Tím je možné model SDXL usměrnit a generovat konzistentnější výsledky.

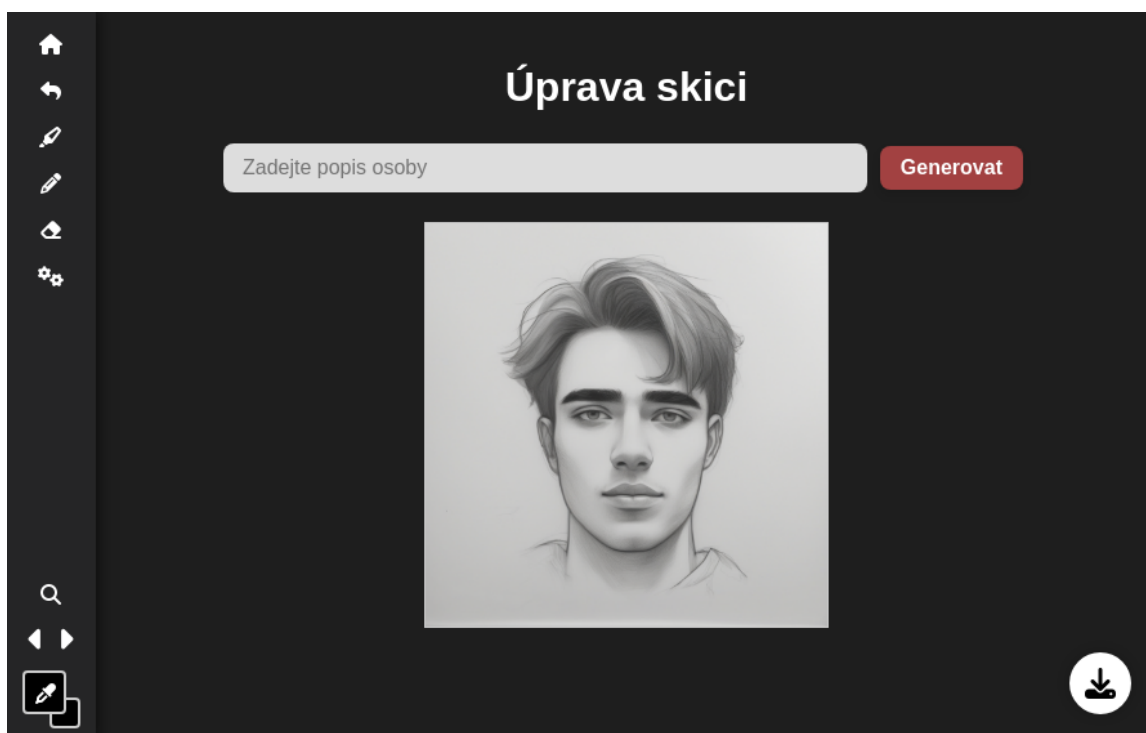
Dalším zjištěním bylo, že Stable Diffusion XL inpainting není vhodný pro změny struktury obličeje (např. posunutí prvků obličeje). Z tohoto důvodu byla aplikace rozšířena o možnost rozložit obličej na jednotlivé prvky pomocí detekce klíčových obličejových bodů a následně s těmito částmi manipulovat. Tato fáze rozšířila možnosti přizpůsobení a umožnila uživateli ovlivnit výsledný vzhled postavy ještě před finální realistickou generací.

V rámci tohoto nástroje může uživatel pomocí jednoduchého uživatelského rozhraní přímo ovládat polohu a rotaci vybraných částí obličeje, konkrétně obočí a očí. Každý z těchto prvků je reprezentován jako dvojice (levý a pravý prvek), se kterou se pracuje synchronně. Pomocí posuvníků lze posunout oči nebo obočí horizontálně i vertikálně, vždy však jako zrcadlový pár, což znamená, že oči nebo obočí se pohybují buď směrem k sobě, nebo od sebe, čímž je zajištěno zachování symetrie obličeje. Stejným způsobem funguje i rotace, kdy se prvky natáčejí vůči své ose tak, aby vznikl přirozeně šikmý výraz, ale přitom si zachovávaly zrcadlovou souměrnost.

Tento způsob manipulace umožňuje uživateli jemně doladit výraz obličeje a vyjádřit emoce či charakterové rysy (např. překvapení nebo zamračení), a to bez nutnosti opakovaného generování nových variant.

#### 4.4.2 Motivace pro generování skici

Celý proces generování identikitu je záměrně rozdělen do tří navazujících fází (generování skici, úprava skici a generování realistického obrázku). Použití skici jako mezikrok mezi textovým popisem a realistickým výstupem se ukázalo jako výhodné, protože umožňuje snadnější úpravy proporcí a rysů obličeje před finálním generováním. Skici jsou méně detailní, proto lépe přizpůsobitelné.



Obrázek 4.2: Ukázka uživatelského rozhraní aplikace.

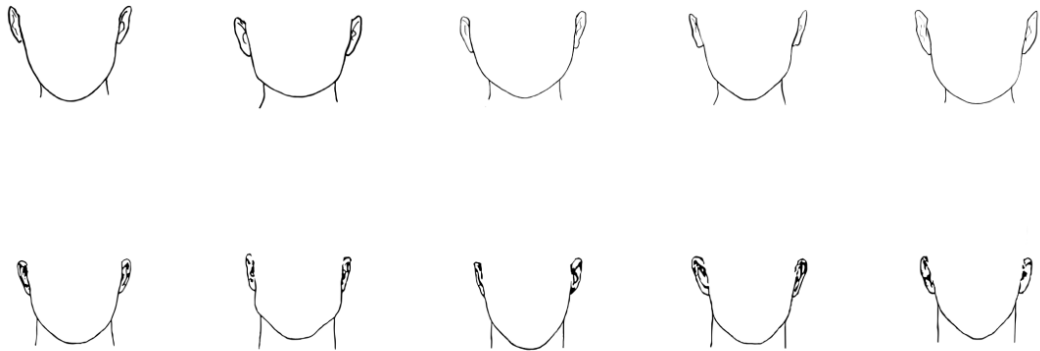
Obrázek 4.2 zobrazuje část aplikace, ve které lze upravovat skicu obličeje a generovat její realistickou podobu. Vlevo se nachází boční panel s nástroji pro úpravu snímku, uprostřed je pracovní plocha s náhledem obličeje a nad ní textové pole sloužící k úpravám nebo generování výstupu na základě zadaného popisu. V pravém dolním rohu se nachází ikona umožňující stažení aktuálního snímku obličeje do zařízení.

#### 4.5 Návrh masek

Pro zvýšení kvality výstupů generovaných modelem Stable Diffusion XL ve spojení s modulem ControlNet bylo nezbytné vytvořit sadu masek, které určují základní tvar obličeje. Tyto masky slouží jako vodítko pro model při generování obličejových rysů a umožňují zachovat proporcčně konzistentní tvar hlavy.

Masky byly navrženy tak, aby reprezentovaly pouze hrubý obrys lidského obličeje, bez vnitřních detailů, jako jsou oči, nos či ústa. Hlavním cílem bylo vytvořit univerzální tvary odpovídající běžným typologiím obličejů, např. kulatý, oválný, hranatý nebo trojúhelníkový tvar. Zároveň bylo nutné z masek odstranit horní část obrysu hlavy, protože při generování vlasů docházelo k problémům, kdy výsledné snímky působily nepřírozně a vlasy se deformovaly.

Při tvorbě masek byly na internetu nejprve vyhledány referenční obrázky, které znázorňovaly typické tvary obličeje. Tyto obrázky byly následně ručně překresleny v grafickém editoru Adobe Photoshop. Důraz byl kladen především na jednoduchost obrysové linie bez jakýchkoli dalších vizuálních prvků. Zhotovené masky byly poté převedeny do formátu s rozlišením  $1024 \times 1024$  pixelů, což odpovídá požadavkům modelu ControlNet určeným pro Stable Diffusion XL.



Obrázek 4.3: Masky reprezentující tvar obličeje.

Na obrázku 4.3 jsou znázorněny tvary dolní části obličeje u žen (horní řada) a mužů (dolní řada). Masky zleva reprezentují jednotlivé typy čelistí: oválný, hranatý, trojúhelníkový, kosočtvercový a srdcový.

## 4.6 Aplikační rozhraní

Aplikační rozhraní tvoří spojení mezi klientskou a serverovou částí systému. Jeho hlavní funkcí je zajištění výměny dat potřebných pro generování a úpravy snímků obličeje.

Přehled koncových bodů je uveden v tabulce 4.1.

Koncový bod	Metoda	Popis
/generate	POST	Generování realistického obličeje.
/generate-sketch	POST	Generování skici obličeje.
/edit	POST	Editace inpaintingem.
/load-masks	GET	Načtení základních tvarů obličeje.
/analyze-face	POST	Rozložení obličeje na části.
/move-face	POST	Transformace částí obličeje.

Tabulka 4.1: Přehled API koncových bodů

## Kapitola 5

# Implementace

Tato kapitola popisuje implementaci aplikace pro generování identikitů na základě textového popisu obličejů. Zaměřuje se na použití modelů Stable Diffusion XL a ControlNet Scribble pro tvorbu obličejů, úpravy vygenerovaných snímků pomocí inpaintingu a manuálních zásahů do skic. Dále se rozebírá rozklad obličejových částí, jejich manipulace a implementace kreslicího rozhraní pro uživatele. Kapitola také popisuje architekturu aplikace, její komunikaci s backendem a výpočetní náročnost. Na závěr se zaměřuje na testování aplikace a její schopnost generovat identikity na základě různých popisů.

### 5.1 Generování obličejů

Pro generování obličejů na základě textového popisu byl zvolen předem natrénovaný model Stable Diffusion XL v kombinaci s ControlNet Scribble. Tato kombinace umožňuje generování obrázků na základě předem nakreslené jednoduché skici, která slouží jako vodítko pro model při vytváření výsledného obrazu. Předem připravené masky, které definují oblasti pro generování, jsou načítány ze složky `masks` na serveru. Při spuštění jsou oba modely pomocí knihovny `Torch` převedeny na grafickou kartu přes rozhraní `CUDA`, což výrazně zrychluje proces generování.

Před procesem generování obrázku bylo nutné přizpůsobit textový popis tak, aby odpovídal požadavkům na výsledné zobrazení. Nejprve je textový vstup přeložen do anglického jazyka, což zajišťuje knihovna `Deep Translator`. Následně je k tomuto textu automaticky přiřazena další textová informace, například zmínka o zobrazení pouze obličejů, vysoké detailnosti, fotorealistickém vzhledu a vzhledu pozadí.

Proces generování je realizován pomocí následujícího úseku kódu:

```
1 image = pipeline(  
2     prompt=prompt,  
3     image=mask_image,  
4     generator=generator,  
5     controlnet_conditioning_scale = 0.7,  
6     width=1024,  
7     height=1024  
8 ).images[0]
```

Zdrojový kód 5.1: Nastavení parametrů pro proces generování.

V kódu 5.1 je využita funkce `pipeline`, která slouží k vygenerování obrázků. Parametr `prompt` obsahuje textový popis požadovaného obrázku. Dále proměnná `image` obsahuje skicu obličeje, která slouží jako vstup pro model ControlNet Scribble. Další parametr `generator` obsahuje náhodný seed, který zajišťuje různorodost výsledků generování. Parametr `controlnet_conditioning_scale` upravuje vliv masky na výsledný obrázek. Konečné rozlišení obrázku je definováno parametry `width` a `height` na hodnotu  $1024 \times 1024$  pixelů.



Obrázek 5.1: Vygenerované snímky obličeje.

Obrázek 5.1 znázorňuje dvojici vygenerovaných snímků obličeje. Levá část představuje základní skicu obličeje vytvořenou na základě textového popisu a předem vytvořené masky reprezentující oválný obličej z části 4.5. Konkrétní popis skici byl „*Mladý muž s krátkými, mírně rozčuchanými vlasy, středně vysokým čelem, výrazným obočím, kulatějšíma očima, rovnějším nosem střední délky, středně plnými rty*“. Pravá část již zachycuje realistickou podobu obličeje, která byla vygenerována s využitím předchozí skici jako referenční masky a doplněného popisu. Popis zněl „*Mladý muž se světlou pleť, tmavě hnědými vlasy, modrýma očima, tmavě hnědým obočím*“. Výsledkem je obraz, který vykazuje vyšší míru realismu a konkrétní vizuální identitu dané osoby.

## 5.2 Editace obličejů

Pro editaci již vygenerovaných obličejů byl využit model Stable Diffusion XL Inpainting, který je určen pro cílené doplňování nebo úpravu částí obrázku definovaných uživatelem.

Vstupní data zahrnují samotný snímek obličeje, masku (kde bílá barva určuje oblasti k úpravě) a textový popis úprav, který je automaticky přeložen do anglického jazyka pomocí knihovny `Deep Translator`.

Kreslení masky je možné na frontendu podobně jako kreslení do obrázku popsané v části 5.3.3. Maska je průhledná a uživatel má k dispozici i nástroj guma pro úpravu masky, která umožňuje mazání nežádoucích částí.

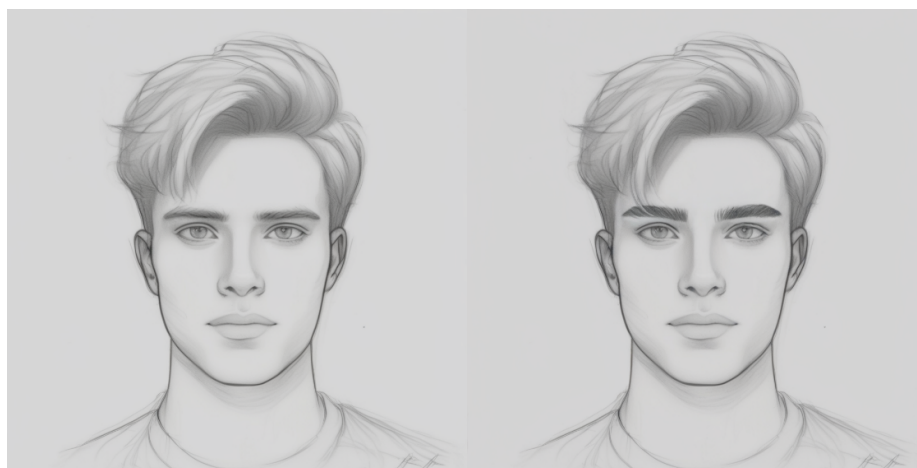
Při editaci skici obličeje byla implementována funkcionalita, která umožňuje úpravy pouze v případě, že je zvolen nástroj pro maskování obrázku. Pokud není tento nástroj aktivní, textové pole slouží pro generování realistické podoby obličeje.

Proces editace je realizován pomocí funkce `pipeline_edit`, která na základě vstupního obrázku, příslušné masky a textového popisu provede úpravu obličeje. Funkce při-

jímá také několik parametrů, které ovlivňují chování modelu při generování. Parametr `guidance_scale` určuje míru, s jakou se má model řídit textovým zadáním, `num_inference_steps` nastavuje počet kroků generování a tím ovlivňuje kvalitu výsledného obrazu, a `strength` definuje intenzitu zásahu modelu do původního obrázku.

Vzhledem k tomu, že výsledky inpaintingu nemusí být vždy zcela předvídatelné a některé části obličeje mohou být vygenerovány nepřirozeně, nebo nemusí odpovídat očekávání uživatele, byla implementována možnost automatického ukládání každého nově vytvořeného snímku do historie úprav. Tato funkce umožňuje přepínat mezi jednotlivými verzemi obličeje, snadno porovnávat předchozí úpravy a případně se vrátit k některé z dřívějších variant, aniž by bylo nutné celý proces opakovat. Přepínání mezi verzemi je možné buď ručně prostřednictvím uživatelského rozhraní, nebo pohodlně pomocí klávesových zkratk `Ctrl+Z` pro krok zpět a `Ctrl+Y` pro krok vpřed.

Stejně jako u modelu pro generování obličejů (viz 5.1) je Stable Diffusion XL Inpainting při spuštění převeden na grafickou kartu pomocí knihovny torch.



Obrázek 5.2: Upravený snímek obličeje.

Obrázek 5.2 zachycuje příklad úpravy obličeje pomocí modelu Stable Diffusion XL Inpainting. Vlevo je původní vygenerovaná skica obličeje, zatímco vpravo je výstup modelu s textovým popisem „*tlusté, výrazné obočí*“. Úprava proběhla pouze v oblasti obočí, definované maskou.

## 5.3 Manuální úprava obličeje

Vzhledem k tomu, že inpainting není spolehlivý při jemných úpravách, jako je například změna polohy prvků, byla přidána možnost ruční editace skici, která slouží jako podklad pro generování obličeje. Uživatel může jednotlivé prvky obličeje posouvat nebo rotovat a dále je možné stávající prvky upravit či nahradit kresbou.

### 5.3.1 Detekce a ořez prvků obličeje

Pomocí knihovny `dlib` je v první fázi na obrázku detekováno 68 obličejových bodů pomocí předtrénovaného modelu `shape_predictor_68_face_landmarks.dat`. Tyto body slouží jako základ pro identifikaci jednotlivých částí obličeje, mezi které patří oči, obočí, ústa, nos a tvar obličeje.

Během testování ořezu jednotlivých částí obličeje bylo zjištěno, že oblast obočí bývá často nevhodně oříznuta. Tento problém vzniká tím, že detekované body obočí se nacházejí pouze na jeho horním okraji, což neumožňuje přesné vymezení celé oblasti obočí. Pro zajištění správného výřezu bylo proto nezbytné doplnit další bod v dolní části obočí, který napomohl ke zpřesnění ořezu.

Podobný problém se vyskytl i při ořezu očí. Původně detekované body vymezovaly pouze oblast oční bulvy, což vedlo k příliš úzkému ořezu, který nezahrnoval oční víčka, ani jejich okolí. Z tohoto důvodu bylo přistoupeno k úpravě, kdy se všechny body z oblasti očí posunuly o 15 pixelů od středu oblasti, čímž došlo k rozšíření ořezu tak, aby zahrnoval celé oko, včetně víček.

Každá partie je ohraničena na základě příslušných bodů a vykreslena jako polygon, který je následně převeden na masku pomocí funkce `cv2.fillPoly`, z knihovny `OpenCV`. Tato maska slouží k výřezu konkrétní oblasti z původního obrázku pomocí bitové operace `cv2.bitwise_and`. Kromě samotného výřezu je uložena i poloha tohoto segmentu v rámci celkového obrázku, což je zásadní pro následné skládání.

### 5.3.2 Posun a rotace prvků obličeje

Pro manipulaci s jednotlivými částmi obličeje byla implementována funkce `posun_oblicej`, která umožňuje provádět jak posun, tak i rotaci partií. Pro dosažení rotace bylo využito funkce `cv2.getRotationMatrix2D`, která generuje rotační transformační matici pro daný úhel a střed rotace. Výsledná matice následně slouží jako vstup do funkce `cv2.warpAffine`, jež aplikuje transformaci zahrnující rotaci a posun. Implementace rotace prvku je znázorněna na 5.2.

```
1 center = (image.shape[1] // 2, image.shape[0] // 2)
2 rotation_matrix = cv2.getRotationMatrix2D(center, rotation_angle, 1)
3 rotated_image = cv2.warpAffine(image, rotation_matrix,
    (image.shape[1], image.shape[0]))
```

Zdrojový kód 5.2: Rotace prvku kolem svého středu o daný úhel.

### 5.3.3 Kreslení do snímku obličeje

Pro realizaci kresby do obrázku byla na straně frontendu implementována interaktivní kreslicí vrstva pomocí HTML prvku `<canvas>`, který je umístěn nad vygenerovaným obrázkem. Kreslení je zajištěno prostřednictvím posluchačů událostí myši, které reagují na její pohyb a stisk tlačítek.

Uživatel má možnost kreslit jak do vygenerované skici obličeje, tak i do realistického vygenerovaného snímku obličeje. Při úpravě skici lze volit mezi černou barvou, nebo nástrojem kapátka, kterým lze vybrat barvu přímo z vygenerovaného obrázku. Kapátko lze aktivovat buď kliknutím na jeho ikonu v uživatelském rozhraní, nebo pomocí klávesové zkratky `CTRL+ALT`.

Při kreslení do realistického snímku obličeje, například pro doplnění tetování nebo mateřských znamének, lze vybírat barvy pouze z předdefinované palety.

Součástí kreslicího režimu je také modální okno, které se zobrazí po kliknutí pravým tlačítkem myši. Uživatel v něm může nastavovat tloušťku čáry. Kromě toho je k dispozici nástroj guma, která umožňuje mazat dříve provedenou kresbu.

Po dokončení kreslení je obsah kreslicí vrstvy sloučen s původním obrázkem buď před odesláním na backend k dalším úpravám nebo před exportem do zařízení. Sloučení je realizováno prostřednictvím pomocného neviditelného prvku `<canvas>`, do kterého se nejprve vykreslí základní obrázek a následně na něj překreslí obsah kreslicí vrstvy.

## 5.4 Implementace frontendu

Frontend zajišťuje vizuální rozhraní, správu nastavení a komunikaci s backendem.

### 5.4.1 Architektura klientské části

Frontendová část aplikace byla realizována s využitím frameworku **Electron**. Aplikace běží jako samostatný desktopový program, jehož hlavní okno načítá jednotlivé HTML stránky odpovídající jednotlivým fázím zpracování dat. Funkcionalita je rozdělena do několika částí, přičemž jednotlivé stránky reprezentují postupné kroky od nastavení parametrů až po finální úpravu vygenerovaného obrázku.

Pro komunikaci mezi hlavním procesem Electronu a uživatelským rozhraním je využito rozhraní IPC (Inter-Process Communication), konkrétně objekty `ipcMain` a `ipcRenderer`. To umožňuje bezpečné předávání parametrů a volání funkcí mezi logikou aplikace a jejím vizuálem. Tato architektura rovněž slouží jako prostředník pro volání HTTP požadavků na server.

Pro zajištění uchování uživatelských nastavení i po zavření a opětovném spuštění aplikace je využita knihovna `electron-store`. Ta umožňuje snadné ukládání a načítání hodnot do lokálního úložiště na straně klienta bez nutnosti použití externí databáze. V lokálním úložišti jsou uloženy hodnoty určující, kolik obrázků se má generovat v jednotlivých fázích zpracování obličeje.

### 5.4.2 Komunikace se serverem

Pro komunikaci s backendem využívá frontend HTTP požadavky odesílané pomocí knihovny `axios`.

Aby bylo možné jednoduše měnit IP adresu a port backendového serveru, aniž by bylo nutné upravovat zdrojový kód, byl přidán konfigurační soubor `config.json`. Tento soubor obsahuje klíč `"backendUrl"`, který definuje adresu serveru, na který jsou odesílány požadavky.

Příklad odeslání požadavku na generování obličeje pomocí knihovny `axios` je uveden níže (5.3).

```
1  const response = await axios.post(`${backendUrl}/generate`, {
2    text: prompt,
3    mask: mask,
4    num_of_repeats: num_of_real_faces
5  });
```

Zdrojový kód 5.3: Odeslání požadavku na generování obličeje.

## 5.5 Výpočetní náročnost

V praktické části práce se objevila zásadní komplikace spojená s vysokou výpočetní náročností generování obrázků pomocí neuronových sítí. Proces byl natolik zatěžující, že běžné osobní počítače nebyly schopny zajistit potřebný výkon, což vedlo k dlouhým prodlevám a doba generování jednoho obrázku trvala několik minut. Tato situace výrazně zpomalovala testování a vývoj aplikace.

Pokusy o optimalizaci, například snížením rozlišení nebo použitím jednodušších modelů, vedly ke zhoršení kvality výstupů bez výrazného zrychlení procesu. Po konzultaci s vedoucím práce byl proto využit výpočetní server fakulty vybavený výkonnějšími grafickými kartami. Díky tomu se zkrátila doba generování obrázků na několik sekund a bylo možné pokračovat v práci bez ztráty kvality.

Při kombinovaném použití modelů jako ControlNet, Stable Diffusion XL a Stable Diffusion XL Inpainting se navíc ukázalo, že samotné načítání a provozování modelů vyžaduje alokaci značného množství grafické paměti.

Při testování na notebooku s grafickou kartou NVIDIA GeForce RTX 3070 s 8 GB VRAM trvalo generování jednoho snímku pomocí modelu Stable Diffusion XL přibližně 15–20 minut. U editace obrázku pomocí inpaintingu se doba generování jednoho snímku pohybovala kolem 1 minuty.

Po přechodu na výpočetní server, vybavený grafickou kartou NVIDIA RTX A5000 s 24 GB VRAM, se výrazně zkrátily časy generování. Zatímco vytvoření celého snímku pomocí Stable Diffusion XL zabralo zhruba 20 sekund, inpainting umožnil rychlou úpravu konkrétní části obrazu během přibližně 5 sekund. Tento výrazný nárůst výkonu umožnil efektivnější testování a vývoj aplikace.

## Kapitola 6

# Testování

Tato část práce se zaměřuje na testování a hodnocení použitelnosti aplikace pro generování identikitů na základě textového popisu. Testování se soustředilo nejen na efektivitu samotného generování, ale i na kvalitu a použitelnost uživatelského rozhraní. Cílem bylo zjistit, jak přesně dokáže aplikace převést textový popis do vizuální podoby osoby a jak intuitivní je samotná práce s nástrojem z pohledu běžného uživatele.

V prvním experimentu byl vytvořen identikit na základě fotografie osoby, na kterou bylo nahlíženo po celou dobu testu. Tento experiment měl za cíl prověřit, jak přesně aplikace generuje identikit za plné kontroly nad zadávanými údaji.

V druhém experimentu byl scénář navržen prostřednictvím prostředníka, kdy jedna osoba pozorovala fotografii a druhá zadávala parametry do aplikace. Tento test měl ověřit, jak aplikace zvládá různé úhly popisování, kdy různí pozorovatelé mohou vnímat detaily osoby jinak.

Ve třetím experimentu měli účastníci omezený čas na prohlédnutí fotografie osoby, po kterém měli z paměti popsat její vzhled.

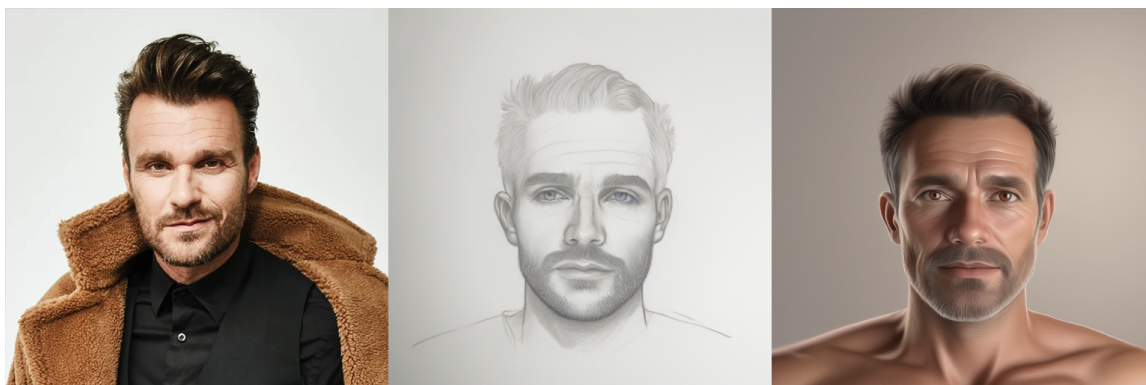
### 6.1 Experiment 1 – Popis i generování jednou osobou

Cílem prvního experimentu bylo ověřit, jak dobře dokáže aplikace vygenerovat identikit na základě textového popisu osoby. Testovaná osoba měla po celou dobu k dispozici fotografii, podle které mohla zadávat popisné údaje přímo do aplikace. Experiment byl navržen tak, aby prověřil přesnost výstupu v ideálních podmínkách, kdy uživatel sám formuluje popis a má plnou kontrolu nad tím, co aplikaci předává.

Test začal výběrem fotografie známé české osobnosti Leoše Mareše, která byla použita jako předloha pro generování identikitu.

V průběhu testu byl použit následující textový popis jako základ pro generování obličeje: „Muž ve věku 40 let s hustými, dozadu učenými vlasy, vyšším a širším čelem, méně výrazným obočím, mírně zapadlými očima, rovnějším nosem střední délky, středně plnými rty a strništěm.“ Pro vytvoření základní sady skic byla zvolena oválná mužská maska. Následně byla z vygenerované nabídky 50 skic obličejů uživatelem vybrána ta, která nejlépe odpovídala osobě z předlohy. Detaily obličeje, jako nos, obočí a účes, byly upravovány převážně pomocí inpaintingu. Vznikla tak skica obličeje, která svými základními rysy připomínala osobu z fotografie.

V závěrečné fázi byla vytvořena realistická podoba identikitu, která však připomínala osobu méně než dříve vytvořená skica. I přesto si identikit ponechal částečnou podobu v základních rysech s předlohou. Výsledný identikit je uveden na obrázku 6.1.



Obrázek 6.1: Reálná fotografie [18] a vygenerovaný identikit Leoše Mareše.

## 6.2 Experiment 2 – Popis svědkem, zadávání úředníkem

Druhý experiment měl simulovat situaci, kdy svědek poskytuje popis osoby, zatímco úřední osoba vytváří identikit.

Testování probíhalo se třemi různými účastníky, kteří se lišili věkem, profesí a schopností vnímat a popisovat vizuální detaily. Účastníci testování byli požádáni o výběr fotografie známé osobnosti jako předlohy, aby bylo později možné využít fotografie osob k vyhodnocení experimentu. Každý ze tří účastníků vystupoval jako svědek a snažil se co nejpřesněji popsat vzhled osoby z fotografie, zatímco další účastník v roli úředníka zadával popis do systému pro generování identikitu.

### Účastník č. 1 – svědek

Prvním účastníkem experimentu byla žena ve věku 51 let, absolventka střední umělecké školy, která se dlouhodobě věnuje kreslení, včetně malby portrétů. Díky své umělecké praxi má tato účastnice vyvinutý cit pro detail a dokázala přesně popisovat jednotlivé rysy obličeje. Jako předlohu si zvolila americkou herečku Scarlett Johansson. Po počátečním popisu vzhledu herečky systém vygeneroval několik základních skic. Jedna z nich účastnici okamžitě zaujala, především díky očím, které podle jejího názoru velmi dobře vystihovaly podobu skutečné osoby. Tato skica se stala základem pro další úpravy, které probíhaly pod jejím vedením.

Účastnice navrhla změny tvaru obočí, vlasů, úst a čelisti. Během procesu systematicky upozorňovala na drobné, ale charakteristické detaily obličeje, například na spadlé koutky rtů. Díky jejím přesným a cíleným popisům vznikla finální skica, která byla hodnocena jako věrně zachycující vzhled předlohy.

Následovalo generování realistické podoby obličeje. Tento krok se ukázal být časově náročnější, než práce se samotnou skicou. Přestože zůstala celková struktura obličeje zachována, realistický obraz zpočátku vykazoval některé odlišnosti. Probíhalo tedy opakované generování s jemnými úpravami textového popisu. Vznikly desítky, možná stovky variant, než se podařilo vytvořit výstup, který odpovídal vizuálním charakteristikám herečky. Vý-

sledná realistická podoba působila věrohodně a celkově dobře vystihovala vzhled herečky. Výsledný identikit je uveden na obrázku 6.2.

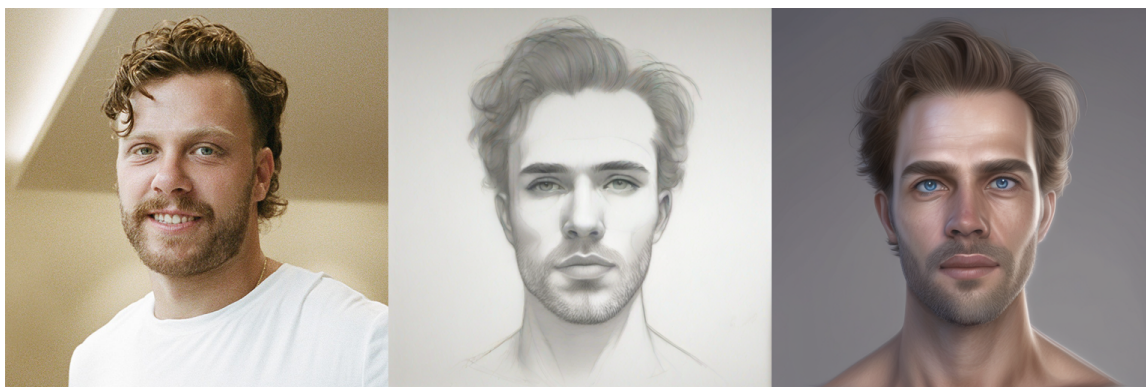


Obrázek 6.2: Reálná fotografie [17] a vygenerovaný identikit Scarlett Johansson.

### Účastník č. 2 – svědek

Druhým účastníkem experimentu byl muž ve věku 23 let, který pracuje v technickém oboru jako projektant dopravních staveb. Jako předlohu si zvolil českého hokejistu Davida Pastrňáka. Po úvodním popisu vzhledu vybrané osoby systém vygeneroval větší množství základních skic, u nichž účastník nenašel shodu. Účastník tak musel postupně porovnávat jednotlivé varianty a vybrat tu, u které byla shoda alespoň částečná. Následovaly úpravy téměř všech částí obličeje.

Po dokončení skici bylo přistoupeno k tvorbě realistické podoby obličeje. I v tomto případě bylo zapotřebí opakované generování v různých variantách s drobnými úpravami popisu. Ačkoli vygenerovaná skica i realistický portrét vykazovaly minimální podobnost s popisovanou osobou, zachovaly si alespoň některé charakteristické rysy obličeje z předlohy. Výsledný identikit je uveden na obrázku 6.3.



Obrázek 6.3: Reálná fotografie [11] a vygenerovaný identikit Davida Pastrňáka.

### Účastník č. 3 – svědek

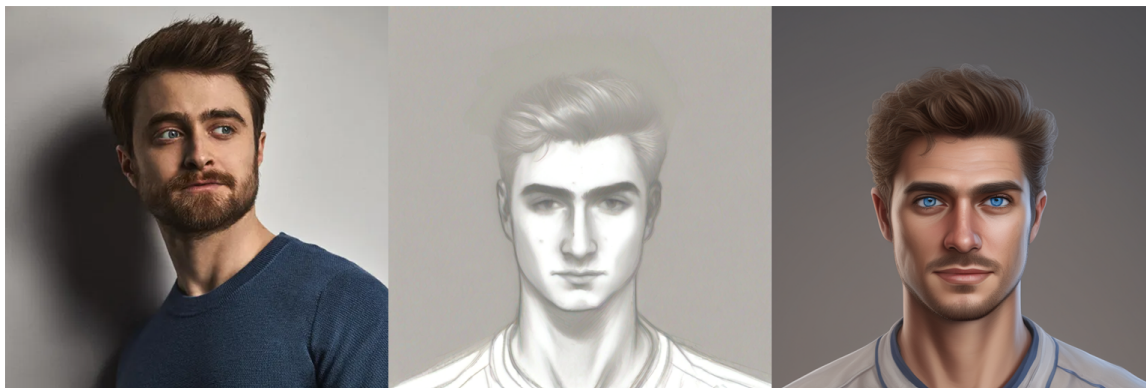
Třetím účastníkem experimentu byl policista ve věku 50 let, s předchozími zkušenostmi v oblasti tvorby identikitů. Jako předlohu si zvolil britského herce Daniela Radcliffa. Po úvodním popisu systém vygeneroval celkem 100 skic obličejů. Dvě z těchto skic účastníkovi

částečně připomínaly vzhled vybrané osoby. Jednu z těchto skic označil jako výchozí pro další úpravy.

Zásadním problémem skici byl nerealisticky vypadající účes, který narušoval celkový dojem. Pokusy o úpravu pomocí inpaintingu nebyly úspěšné, a tak byly vlasy nejprve ručně přemalovány a až poté doladěny pomocí inpaintingu. Dále bylo potřeba upravit oči, obočí, rozšířit nos a zúžit čelist, jelikož ji původní skica zobrazovala příliš masivní. Po provedení těchto úprav začala výsledná skica již částečně připomínat Daniela Radcliffa a zachycovala některé z jeho charakteristických rysů.

Účastník prokázal dobrou schopnost popisovat základní rysy obličeje, zejména jejich proporce, natočení a celkové rozmístění. Oproti první účastnici nevěnoval takovou pozornost detailům.

Po dokončení skici se přistoupilo ke generování realistické podoby obličeje. Výsledná verze sice vykazovala určité rysy, které byly inspirovány původní skicou a popisem, nicméně celková vizuální podobnost s Danielem Radcliffem byla velmi nízká. Navzdory tomu byly zachovány některé základní znaky obličeje, které vykazují alespoň částečnou vizuální podobnost. Výsledný identikit je uveden na obrázku 6.4.



Obrázek 6.4: Reálná fotografie [25] a vygenerovaný identikit Daniela Radcliffa.

### 6.3 Experiment 3 – Popis z paměti

Ve třetím experimentu měli účastníci omezený čas (2 minuty) na prohlížení fotografie osoby, po kterém dostali za úkol popsat z paměti její vzhled. Tento test měl za cíl zjistit, jak kvalitně dokáže aplikace generovat identikit na základě omezených informací, které účastník získá během krátkého pohledu na fotografii. Účastníci zůstali stejní jako v předchozím experimentu.

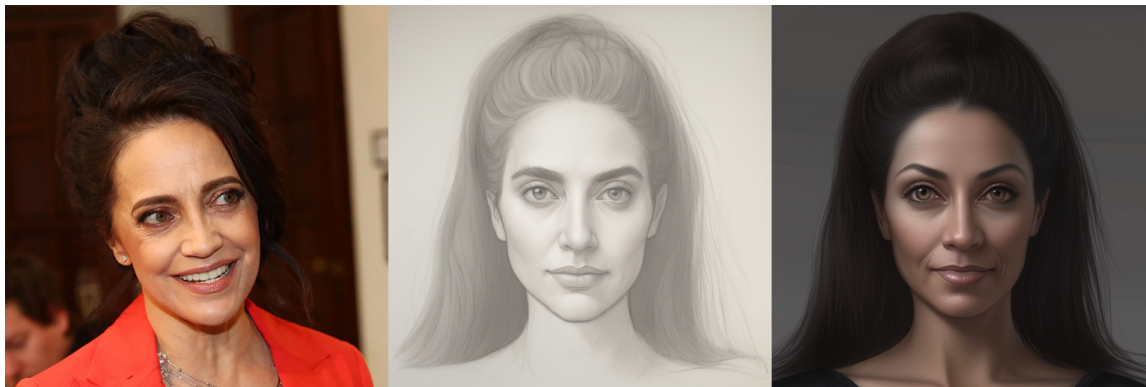
#### Účastník 1 – svědek

Jako předloha pro generování identikitu byla vybrána známá česká zpěvačka Lucie Bílá. Díky své vysoké vizuální představitosti a dobré vizuální paměti si účastnice dokázala velmi dobře vybavit vzhled osoby a popsat základní části obličeje.

Po prvotním popisu bylo vygenerováno celkem 100 skic. Účastnice si z této nabídky vybrala skicu, která vystihovala výrazné oči Lucie Bílé. Poté začala samotná úprava prvků obličeje. Poměrně náročnou částí bylo vygenerování tvaru obočí, které je u Lucie Bílé výrazně zakulacené. Tento problém byl vyřešen rotací obočí, která byla následně kombinována

s inpaintingem. Popis osoby byl velmi precizní, což vedlo k vytvoření skici, která vykazuje podobnost se základními rysy Lucie Bílé.

Po dokončení skici následovalo generování realistické podoby obličeje. Přestože skica zcela neodpovídala skutečné podobě, generovaný realistický identikit byl velmi kvalitní a poměrně dobře vystihoval vzhled zpěvačky. Výsledný identikit je uveden na obrázku 6.5.



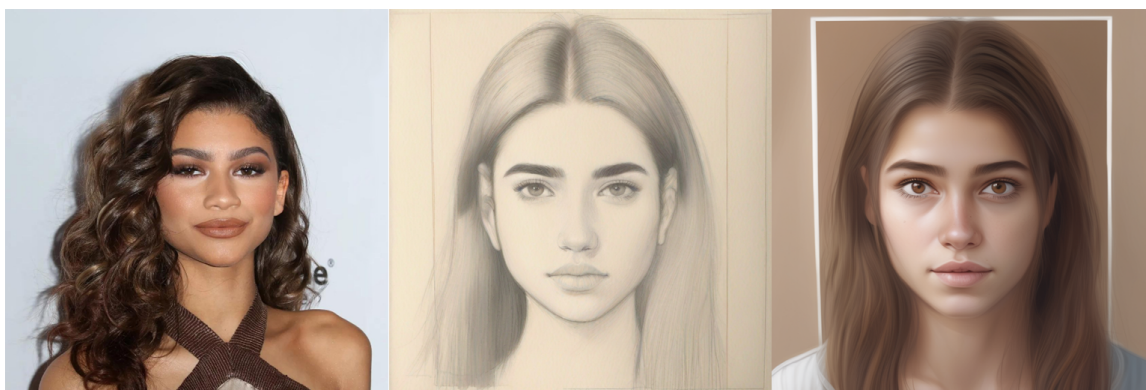
Obrázek 6.5: Reálná fotografie [31] a vygenerovaný identikit Lucie Bílé.

## Účastník 2 – svědek

Jako předloha pro generování identikitu byla účastníkem vybrána známá americká herečka Zendaya. Po úvodním popisu vzhledu osoby byla systémem vygenerována první sada skic. Dle vyjádření účastníka neodpovídala ani jedna z vygenerovaných skic zvolené předloze. Účastník proto opakovaně upravoval textový vstup s cílem lépe vystihnout hereččiny rysy, přičemž byly generovány další varianty skic. Přesto se nepodařilo získat výstup, který by se Zendaye alespoň částečně podobal.

Následovala fáze úprav vybrané skici. Pomocí dostupných nástrojů byly upravovány jednotlivé prvky obličeje. Ani přes tyto zásahy se však nedařilo zachytit konkrétní rysy herečky, a to hlavně z důvodu nejistoty účastníka při popisu jednotlivých částí obličeje.

V závěrečné fázi byl ze skici vygenerován realistický identikit. Ani tento výstup neodpovídal podobě Zendayi. Výsledný identikit je uveden na obrázku 6.6.

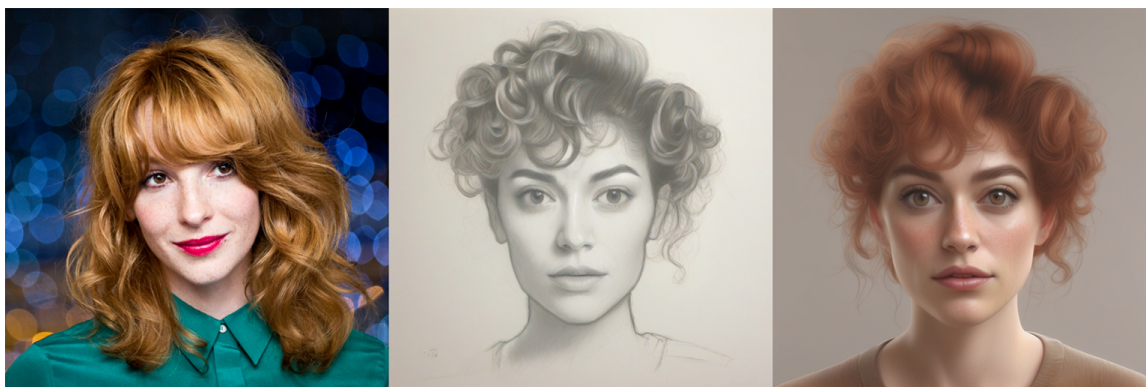


Obrázek 6.6: Reálná fotografie [4] a vygenerovaný identikit Zandayi.

### Účastník 3 – svědek

Třetí účastník si zvolil herečku Vicu Kerekes. Po uplynutí dvou minut, během nichž měl možnost si herečku prohlédnout, poskytl popis, ve kterém vyzdvihl zejména její kulaté oči. Na základě těchto informací byla vygenerována první sada skic, z níž si účastník vybral variantu s výrazně kulatými očima.

Následovala úprava jednotlivých prvků skici, především očí a obočí. I přes snahu o doladění však nebyla výsledná skica příliš přesvědčivá a postrádala klíčové rysy obličeje. Problémem bylo především, že účastník nebyl schopen detailního popisu jednotlivých částí obličeje. Realistická podoba, která byla vygenerována na základě této skici, nicméně dokázala lépe zachytit specifické rysy Vicy Kerekes, zejména pihatou pleť. Výsledný identikit je uveden na obrázku 6.7.



Obrázek 6.7: Reálná fotografie [15] a vygenerovaný identikit Vicy Kerekes.

## 6.4 Testování uživatelského rozhraní

Pro ověření použitelnosti aplikace a intuitivnosti jejího ovládání bylo provedeno testování uživatelského rozhraní. Cílem tohoto testování bylo prověřit všechny dostupné funkcionality a zjistit případné nedostatky v interakci s uživatelem.

Testování se zúčastnili dva dobrovolníci, kteří postupně prošli celým procesem generování identikitu. Účastníci byli obeznámeni s cílem testování a měli za úkol hodnotit ovladatelnost, srozumitelnost jednotlivých kroků a přehlednost celé aplikace.

### Zpětná vazba uživatelů

Na základě zpětné vazby od účastníků byly do aplikace zavedeny konkrétní úpravy, které zlepšují uživatelský komfort a předcházejí nechtěným chybám při práci s obrazem. Byla implementována funkce pro zvětšení a zmenšení obrázku, což umožňuje přesněji pracovat s jeho detaily, například při kreslení. Dále byla zavedena potvrzovací okna při návratu na úvodní stránku nebo o jednu stránku zpět. Tato funkce slouží jako ochrana proti nechtěné ztrátě rozpracovaného obrázku. Rovněž bylo přidáno potvrzovací okno při generování realistického portrétu, které uživatele upozorní, že aktuálně není aktivována funkce maskování. Bez této funkce by došlo ke spuštění generování místo provádění úpravy, což bylo vyhodnoceno jako častý zdroj nedorozumění a chyb.

### Možnosti budoucího rozšíření

Testování aplikace přineslo nejen cennou zpětnou vazbu k aktuálním funkcím, ale také inspirovalo k dalšímu možnému vylepšení, které by mohlo přispět k výraznému zlepšení

funkcionality. Účastníci projevili zájem o zavedení funkce razítka, která by umožňovala kopírovat vybrané části obrázku na jiné místo. Tato možnost by byla obzvláště užitečná pro úpravu jemných detailů nebo odstraňování nežádoucích prvků, které se při generování obrázku mohou občas objevit.

## 6.5 Výsledky testování

Testování aplikace pro generování identikitů odhalilo několik zásadních poznatků o jejích schopnostech i omezeních. Aplikace dokázala poměrně přesně generovat základní rysy obličeje, pokud byl k dispozici dostatečně podrobný a přesný textový popis.

Na základě testování bylo zjištěno, že jedním z hlavních problémů je samotná jazyková interpretace vzhledu. Mnozí účastníci, zvláště ti bez odborné zkušenosti, nedokázali přesně jazykově vystihnout tvarové detaily obličeje. I v případech, kdy byl popis dílčí části velmi dobrý, model SDXL občas vygeneroval výstup opačný, nebo velmi nepřesný. Ukazuje se tak, že i samotný model má své limity a nedokáže vždy správně interpretovat textový vstup. Tento problém by mohlo částečně vyřešit nasazení výkonnějších generativních modelů, jako je například DALL-E nebo novější verze Stable Diffusion 3.0 či 3.5, které zvládají lépe zpracovat přirozený jazyk a jsou trénovaná na větším množství dat.

Dále bylo zjištěno, že při popisu jednotlivých částí obličeje je výhodné zahrnout více atributů. Například při popisu řídkého obočí bylo efektivnější použít detailnější popis jako „jemné, úzké, nevýrazné, řídké obočí“. Takový popis poskytuje detailnější informaci a pomáhá vytvořit přesnější výstup.

Z testování dále vyplynulo, že ideální strategií pro dosažení věrohodného identikitu je vytvoření co největšího počtu základních skic podle popisu. Ačkoliv je tento postup časově náročný, významně zvyšuje šanci, že se alespoň jedna ze skic bude blížit reálné podobě hledané osoby.

Provedené experimenty potvrdily, že si lidé spíše pamatují tváře jako celek než jejich jednotlivé části, a svědci tak často nedokážou detailně popsat konkrétní rysy obličeje. Naopak některé osoby s uměleckým nadáním si určité detaily dokážou částečně zapamatovat.

Z výše uvedeného je zřejmé, že ačkoliv je aplikace funkční a užitečná při tvorbě základního identikitu, její přesnost je silně závislá na kognitivních a interpretačních schopnostech svědka.

# Kapitola 7

## Závěr

Cílem této bakalářské práce bylo navrhnout a vyvinout funkční aplikaci pro generování portrétních identikitů za pomoci moderních technik strojového učení. V rámci výzkumu byla vyvinuta aplikace, která využívá model Stable Diffusion XL pro generování vizuálních podob osob na základě textových popisů.

Aplikace je schopná generovat identikit v dostatečné kvalitě, pokud byl k dispozici detailní a správný textový popis. Problémy nastaly při interpretaci složitějších detailů vzhledu, kdy model, i přes pokročilé techniky, občas vytvořil nepřesné nebo zcela opačné výstupy. Tato zjištění ukazují, že i moderní generativní modely, jako je Stable Diffusion XL, mají své limity, přičemž další vylepšení by bylo možné díky využití novějších modelů, jako je DALL-E nebo novější verze Stable Diffusion.

Při tvorbě identikitu narážíme též na kognitivní psychologické bariéry spojené s tím, jak si lidé pamatují lidské tváře. Lidský mozek vnímá tváře spíše jako celek a není schopen věrohodně interpretovat jednotlivé komponenty obličeje. Tento faktor přináší zásadní problémy při generování identikitu. Testováním bylo zjištěno, že ideálním přístupem pro dosažení věrohodného identikitu spočívá v generování více skic na základě jednoho popisu. Tento proces, ačkoli časově náročný, zvyšuje pravděpodobnost, že výsledný identikit bude odpovídat realitě.

V aplikaci byla implementována možnost manuálních úprav jednotlivých částí obličeje z důvodu nižší spolehlivosti při generování jemných detailů prostřednictvím generativních modelů. Tato funkce umožnila přizpůsobit části obličeje tak, aby odpovídaly požadavkům svědka.

Testování aplikace ukázalo, že generování obrázků pomocí neuronových sítí je výpočetně náročné a vedlo k dlouhým prodlevám na běžných počítačích. Z tohoto důvodu bylo přistoupeno k využití výpočetního serveru, kdy se doba potřebná pro generování řádově zkrátala.

V rámci bakalářské práce se podařilo naplnit stanovené cíle. Byla úspěšně navržena a implementována aplikace v jazyce Python, která umožňuje generovat identikit na základě textového popisu v českém jazyce s využitím moderních neuronových sítí. Byl prozkoumán proces tvorby identikitů z pohledu kriminalistiky a antropologie a zhodnoceny dostupné technologie pro generování obličejů. Součástí řešení byla také integrace manuálních úprav obličeje pro zvýšení přesnosti výstupu. Aplikace byla otestována s uživateli, čímž byla ověřena její funkčnost i praktická využitelnost.

# Literatura

- [1] BANJARA, B. *Unleashing Generative AI with VAEs, GANs, and Transformers* online. Dostupné z: <https://www.analyticsvidhya.com/blog/2023/07/generative-ai-with-vaes-gans-transformers/>. Poslední úprava: 11. srpna 2023. [vid. 2025-5-7].
- [2] BERGMANN, D. a STRYKER, C. *What is a variational autoencoder?* online. Dostupné z: <https://www.ibm.com/think/topics/variational-autoencoder>. Poslední úprava: 12. června 2024. [vid. 2024-12-9].
- [3] BONDARENKO, S. *Aktivační funkce: skoková, lineární, sigmoidní, ReLU a Tanh* online. Dostupné z: <https://robotdreams.cz/blog/122-aktivacni-funkce-skokova-linearni-sigmoidni-relu-a-tanh>. Poslední úprava: 2025. [vid. 2024-12-8].
- [4] BRIONES, I. *Zendaya Reveals the Important Reason She Returned to Disney Channel* online. Dostupné z: <https://www.teenvogue.com/story/zendaya-disney-channel-representation>. Poslední úprava: 11. dubna 2016. [vid. 2025-5-9].
- [5] BUCKALEW, A. *Supervised vs Unsupervised Learning Explained* online. Dostupné z: <https://www.seldon.io/supervised-vs-unsupervised-learning-explained-2/>. Poslední úprava: 9. března 2025. [vid. 2025-3-23].
- [6] BUETTGENBACH, M. H. *Explain like I'm five: Artificial neurons* online. Dostupné z: <https://towardsdatascience.com/explain-like-im-five-artificial-neurons-b7c475b56189>. Poslední úprava: 8. listopadu 2021. [vid. 2024-11-18].
- [7] CHAUDHARY, K. *What are Autoregressive Generative Models* online. Dostupné z: <https://drops of ai.com/what-are-autoregressive-generative-models/>. Poslední úprava: 21. srpna 2024. [vid. 2024-1-9].
- [8] CHEN, S.-Y.; SU, W.; GAO, L.; XIA, S. a FU, H. DeepFaceDrawing: deep generation of face images from sketches. *ACM Transactions on Graphics (TOG)*, 2020, sv. 39, č. 4, s. 72:1–72:16. Dostupné z: <https://doi.org/10.1145/3386569.3392386>.
- [9] DAVE BERGMANN, C. S. *What are diffusion models?* online. Dostupné z: <https://www.ibm.com/think/topics/diffusion-models>. Poslední úprava: 21. srpna 2024. [vid. 2024-12-9].
- [10] DAVIS, B. *IQ Biometrix FACES 4.0* online. Dostupné z: <https://www.policemag.com/technology/article/15349574/iq-biometrix-faces-40>. Poslední úprava: 30. dubna 2004. [vid. 2025-4-5].

- [11] DEBNÁROVÁ, N. *David Pastrňák: Nohama pevně na ledě* online. Dostupné z: <https://www.vogue.cz/clanek/vogue-daily/natalie-debnarova/david-pastrnak-nohama-pevne-na-lede>. Poslední úprava: 30. srpna 2024. [vid. 2025-5-8].
- [12] ELHARROUSS, O.; ALMAADEED, N.; AL-MÁADEED, S. a AKBARI, Y. Image inpainting: A review. *Neural Processing Letters*, 2020, sv. 51, s. 2007–2028. Dostupné z: <https://doi.org/10.1007/s11063-019-10163-0>.
- [13] GUI, J.; SUN, Z.; WEN, Y.; TAO, D. a YE, J. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 2023, sv. 35, č. 4, s. 3313–3332. Dostupné z: <https://doi.org/10.1109/TKDE.2021.3130191>.
- [14] HARDESTY, L. *Neural networks* online. Dostupné z: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>. Poslední úprava: 14. dubna 2017. [vid. 2024-10-25].
- [15] HEROUT, V. *České celebrity* online. Dostupné z: <https://www.vojtaherout.com/cz/portfolio/ceske-celebrity/>. Poslední úprava: 2025. [vid. 2025-5-9].
- [16] HOURIHAN, K. L.; BENJAMIN, A. S. a LIU, X. A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition*, 2012, sv. 1, č. 3, s. 158–162. Dostupné z: <https://doi.org/10.1016/j.jarmac.2012.06.004>.
- [17] KOMONIBO, I. *Scarlett Johansson Is Going To Battle With Disney Over Black Widow* online. Dostupné z: <https://www.refinery29.com/en-us/2021/07/10605681/scarlett-johansson-disney-black-widow-lawsuit-streaming>. Poslední úprava: 29. července 2021. [vid. 2025-5-8].
- [18] MAREŠ, M. *Milion followerů! Jsem obchodník, tenhle status mi může pomoci, říká Leoš Mareš* online. Dostupné z: <https://forbes.cz/milion-followeru-jsem-obchodnik-pomuze-to-i-charite-rika-leos-mares>. Poslední úprava: 3. června 2020. [vid. 2025-5-5].
- [19] MICEK, G. *FACES 4.0 Professional* online. Dostupné z: <https://facialcomposites.com/products/faces-4-0-pro-software-download>. Poslední úprava: 2025. [vid. 2025-4-5].
- [20] MUSIL, J.; SUCHÁNEK, J. a KONRÁD, Z. *Kriminalistika*. C. H. Beck, 2004. ISBN 80-7179-878-9.
- [21] NOBLE, J. *What is an autoregressive model?* online. Dostupné z: <https://www.ibm.com/think/topics/autoregressive-model>. Poslední úprava: 12. června 2024. [vid. 2025-5-7].
- [22] O’CONNOR, R. *How DALL-E 2 Actually Works* online. Dostupné z: <https://www.assemblyai.com/blog/how-dall-e-2-actually-works>. Poslední úprava: 29. září 2023. [vid. 2025-3-24].
- [23] PORADA, V. *Kriminalistika. Technické, forenzní a kybernetické aspekty*. Aleš Čeněk, 2016. ISBN 978-80-7380-589-0.

- [24] PRA, M. D. *Generative Adversarial Networks* online. Dostupné z: <https://medium.com/@marcodelpra/generative-adversarial-networks-dba10e1b4424>. Poslední úprava: 30. října 2023. [vid. 2024-12-9].
- [25] RITMAN, A. *Daniel Radcliffe on Going ‘Weird’ in His Post-Potter Career, Picking Roles “On the Basis of What Seems Fun”* online. Dostupné z: <https://www.hollywoodreporter.com/movies/movie-features/daniel-radcliffe-weird-career-harry-potter-yankovic-1235252666/>. Poslední úprava: 1. listopadu 2022. [vid. 2025-5-8].
- [26] ROMBACH, R.; BLATTMANN, A.; LORENZ, D.; ESSER, P. a OMMER, B. High-Resolution Image Synthesis with Latent Diffusion Models. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2022, s. 10674–10685. ISBN 978-1-6654-6946-3.
- [27] SAHARIA, C.; CHAN, W.; SAXENA, S.; LI, L. a WHANG, J. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2022, s. 36479 – 36494. ISBN 978-1-7138-7108-8.
- [28] SANDNER, D. *Flux Tools: New Outpainting, Redux IP Adapter Solution, and Controlnet LoRAs?* online. Dostupné z: <https://sandner.art/flux-tools-new-outpainting-redux-ip-adapter-solution-and-controlnet-loras/>. Poslední úprava: 24. listopadu 2024. [vid. 2025-4-3].
- [29] STRAUS, J. *Kriminalistická technika*. Aleš Čeněk, 2012. ISBN 978-80-7380-409-1.
- [30] TANAKA, J. W. a FARAH, M. J. Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A*, 1993, sv. 46, č. 2, s. 225–245. Dostupné z: <https://doi.org/10.1080/14640749308401045>.
- [31] ULMANOVÁ, L. *Kvíz: Lucie Bílá slaví narozeniny! Znáte její slavné hity? A víte, kolikrát za život se vdala?* online. Dostupné z: <https://www.prozeny.cz/clanek/lucie-bila-slavi-54-narozeny-co-vite-o-zname-zpevacke-63261>. Poslední úprava: 7. dubna 2024. [vid. 2025-5-8].
- [32] VALKOVIČ, P. *Lekce 2 - Neuronové sítě - Perceptron* online. Dostupné z: <https://www.itnetwork.cz/python/neuronove-site/pokrocile/neuronove-site-perceptron>. Poslední úprava: 2025. [vid. 2024-11-24].
- [33] WENG, L. *From Autoencoder to Beta-VAE* online. Dostupné z: <https://lilianweng.github.io/posts/2018-08-12-vae/#vq-vae-and-vq-vae-2>. Poslední úprava: 12. srpna 2018. [vid. 2025-5-7].
- [34] ZHANG, C.; ZHANG, C.; ZHANG, M.; KWEON, I. S. a KIM, J. *Text-to-image Diffusion Models in Generative AI: A Survey*. 2024. Dostupné z: <https://arxiv.org/abs/2303.07909>. [vid. 2025-5-7].
- [35] ZHANG, L.; RAO, A. a AGRAWALA, M. Adding Conditional Control to Text-to-Image Diffusion Models. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2023, s. 3813–3824. ISBN 979-8-3503-0718-4.

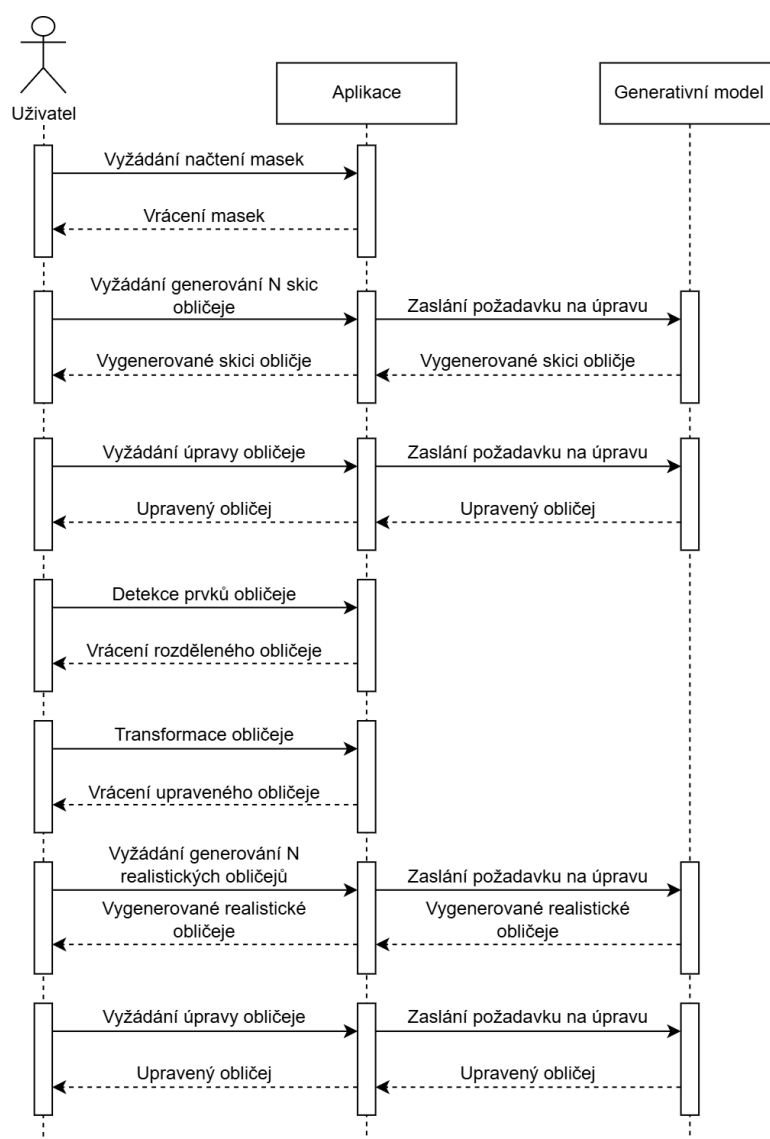
# Příloha A

## Obsah paměťového média

- `README.md` – Popis projektu a návod ke spuštění.
- `src/` – Zdrojové soubory aplikace:
  - `frontend/` – Klientská část.
  - `backend/` – Serverová část.
- `latex/` – Zdrojové soubory bakalářské práce.
- `xfurik00.pdf` – Finální verze práce ve formátu PDF.
- `video.mp4` – Ukázka funkčnosti aplikace.

## Příloha B

# Sekvenční diagram



Obrázek B.1: Sekvenční diagram aplikace.