

# GYMNASIUM JANA KEPLERA

Parléřova 2/118, 169 00 Praha 6



## SimonPose

### Maturitní práce

Autor: Šimon Brávek

Třída: 4.B

Školní rok: 2025/2026

Předmět: Informatika

Vedoucí práce: Jiří Matas

Praha, 2026





# Gymnázium Jana Keplera

Kabinet informatiky

## ZADÁNÍ MATURITNÍ PRÁCE

---

- *Student:* Šimon Brávek
  - *Třída:* 4.B
  - *Školní rok:* 2025 / 2026
  - *Vedoucí práce:* Jiří Matas
  - *Název práce:* Vylepšení odhadu pózy člověka z jednoho vstupního snímku
  - *URL repozitáře:* <https://github.com/simonbravek/pose-detection-project>
- 

### Pokyny pro vypracování:

Cílem práce je navrhnout, implementovat a experimentálně ověřit metodu pro **přesnější odhad lidské pózy** z jednoho RGB snímku.

Student využije již existující model **DensePose** pro získání hustého korespondenčního mapování povrchu těla a naváže na něj vlastní optimalizační nebo post-processingovou metodu s cílem:

- snížit chyby v odhadu pozice a orientace klíčových kloubů,
- případně zvýšit robustnost vůči šumu a odlišnostem v pozici či osvětlení vstupního snímku.

### Součástí práce je:

- rešerše současných přístupů k odhadu pózy,
- implementace vlastního řešení v prostředí Python (TensorFlow),
- návrh metrik pro hodnocení přesnosti a srovnání s výchozím stavem,
- experimentální vyhodnocení na vybraném datasetu,
- prezentace dosažených výsledků a postupu práce formou závěrečné zprávy a ukázky funkčního systému.

### Výstupy práce:

- zdokumentovaný software v repozitáři,
- technická zpráva shrnující postup, metodiku a dosažené výsledky,
- prezentace práce a její obhajoba.

---

Podpis vedoucího práce:

A handwritten signature in blue ink, appearing to read "J. Matas". The signature is written in a cursive, flowing style.

## **Prohlášení**

Prohlašuji, že jsem svou práci vypracoval samostatně a použil jsem pouze prameny a literaturu uvedené v seznamu bibliografických záznamů. Nemám žádné námitky proti zpřístupňování této práce v souladu se zákonem č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších předpisů.

V Praze dne 27. února 2026

Šimon Brávek



## Poděkování

Děkuji Jiřímu Matasovi za skvělou příležitost strávit měsíc v létě na katedře strojového učení a být součástí soudobého výzkumu a za jeho skvělé nápady v tomto oboru. Bezpochyby tato zkušenost formovala mé rozhodnutí studovat obor AI na MFF UK a propůjčila mi vášeň k nejnovějším technologiím. Děkuji také Matěji Suchánkovi jako člověku, co mě provedl tou nejtěžší asimilací s prací s grafickými kartami a modely jež jsem používal. Děkuji také lidem z katedry, jež ke mě byli přátelští a v neposlední řadě Igoru Vujovičovi za rozvíjení informatiky na našem gymnáziu i přes mnohé překážku.





## **Abstrakt**

Jde o explorativní analýzu s cílem přinést metody pro zlepšení human pose estimation.

## **Klíčová slova**

počítačové vidění, odhad lidské pozice, DensePose, SMPL/SMPL-X,  
rekonstrukce z 2D do 3D, hluboké učení, vylepšení modelu

## **Abstract**

## **Keywords**

computer-vision, human-pose-estimation, DensePose, SMPL/SMPL-X,  
2D-to-3D reconstruction, deep-learning, model improvement



# Obsah

<b>Úvod</b>	<b>3</b>
0.1 Analýza postavy v digitálním světě . . . . .	3
0.2 Problematika . . . . .	3
0.3 Cíl práce . . . . .	3
<b>1 Teoretická část</b>	<b>5</b>
1.1 Základní teorie . . . . .	5
1.1.1 Co je to model? . . . . .	5
1.1.2 Druhy detekcí . . . . .	5
1.1.3 Odhad 2D a jeho limitace . . . . .	5
1.1.4 Od 2D bodů k 3D objemu . . . . .	6
1.2 Skinned Multi-Person Linear Model (SMPL) . . . . .	6
1.2.1 Výpočet . . . . .	6
1.2.2 Výhody . . . . .	8
1.3 DensePose . . . . .	8
1.3.1 DensePose v1: mapování pomocí (I, U, V) . . . . .	9
1.3.2 DensePose v2: Continuous Surface Embeddings (CSE) . . . . .	9
1.3.3 Silné a slabé stránky . . . . .	10
<b>2 Explorativní část</b>	<b>13</b>
2.1 Vzdálenost bodu na těle . . . . .	13
2.1.1 Návrh metody . . . . .	13
2.1.2 Implementace . . . . .	14
2.1.3 Výsledky . . . . .	14
2.2 Euklidovská vzdálenost . . . . .	14
2.2.1 Návrh metody . . . . .	14
2.2.2 Implementace . . . . .	14
2.2.3 Výsledky . . . . .	14
2.3 Zaměření na přesnost . . . . .	15
2.3.1 Návrh metody . . . . .	15
2.3.2 Implementace . . . . .	15
2.3.3 Výsledky . . . . .	15
<b>3 Technická dokumentace</b>	<b>17</b>
3.1 Ukázka sekce . . . . .	17
3.1.1 A jedné podsekce . . . . .	17
3.2 A další sekce . . . . .	18
<b>Závěr</b>	<b>21</b>
<b>Seznam použité literatury</b>	<b>23</b>
<b>Seznam obrázků</b>	<b>25</b>
<b>Seznam tabulek</b>	<b>26</b>



# Úvod

## 0.1 Analýza postavy v digitálním světě

V současné informatice a počítačovém vidění (Computer Vision) představuje automatické porozumění lidského pohybu jednu z největších výzev. Nejde již o prostou detekci zda se v obrázku nachází člověk, či kolik jich je, ale o snahu přenést lidskou biomechaniku do digitálního prostoru. Schopnost přesně interpretovat lidskou pózu z běžného 2D obrazu (např. z mobilního telefonu) otevírá dveře aplikacím, které byly dříve nemyslitelné bez drahých studiových systémů pro snímání pohybu (Motion Capture).

Tyto technologie se dnes nachází již v mnoha podobách a můžou tak pomoci v celé řadě aplikací. Dnes již můžeme najít rozpoznání člověka a obličeje v mobilní aplikaci galerie, kde napomáhá chytrému třídění fotek nebo v bezpečnostních systémech pro detekci vetřelce. Detekce konkrétní pozice, tedy kde se nachází jednotlivé klouby, může zase pomoci zastavit automatický vozík před kolizí s člověkem nebo zavolat pomoc pokud uvidí nehybně ležícího chodce.

Kromě přesnosti těchto metod se vědci čím dál více zaměřují také na rychlost. Mnohé aplikace totiž potřebují živou analýzu videa a to často i z několika kamer najednou. Tak tomu je u autonomního řízení automobilů, které se ukazuje být velkým tahounem tohoto odvětví.

Právě rozmanitost této disciplíny, její užitečnost a také možnost setkat se s ní do hloubky na FEL ČVUT mě vedlo k její volbě jako tématu mé maturitní práce.

## 0.2 Problematika

Současné SOTA (State-of-the-Art) modely sice dosahují vynikajících výsledků na laboratorních datasech, ale často selhávají v reálných podmínkách (in-the-wild). Mezi kritické faktory patří:

- **Zákryty (Occlusions):** Části těla zakryté předměty nebo jinými částmi těla.
- **Perspektivní zkreslení:** Extrémní úhly kamery, které deformují vizuální proporce těla.
- **Vizuální šum:** Volné oblečení nebo špatné světelné podmínky.
- **Propletení více těl:** Situace v davu, kde není jasné, komu patří jaká část těla a jak je správně rozdělit.

Právě kombinace modelu SMPL s technologií DensePose (která mapuje pixely přímo na povrch těla) nabízí unikátní cestu, jak tyto problémy řešit skrze hustou korespondenci dat.

## 0.3 Cíl práce

Cílem této maturitní práce není vytvořit nový, revoluční model, který by překonal stávající vědecké rekordy. Ambicí je metodický průzkum možností, jak stávající proces fittingu (pasování) modelu SMPL do dat z DensePose zpřesnit a učinit jej odolnějším (robustnějším).

Zvolená metodika se opírá o princip Fail-Fast:

- Rychlá formulace hypotéz o vylepšení optimalizačního procesu.
- Implementace prototypů a jejich testování na hraničních případech.
- Analýza selhání jakožto hlavního zdroje poznání.

Výsledkem práce je ucelený přehled vyzkoušených metod, jejich kritické zhodnocení a dokumentace slepých i perspektivních uliček, které mohou sloužit jako inspirace pro další vývoj v oblasti monokulární 3D rekonstrukce člověka.

# 1. Teoretická část

## 1.1 Základní teorie

### 1.1.1 Co je to model?

V této práci se často budu věnovat programům, které využívají metodu strojového učení k detekci a odhadu pozice člověka v obrázku, který nazýváme model. Tento program má mnoho parametrů na základě nichž přetváří vstup na výstup. Nejdříve projde procesem, které nazýváme trénování. Během toho se jako vstup použije obrázek a program k němu vygeneruje souřadnice, kde odhaduje klouby, na základě prvního nastavení parametrů, neboli inicializace. Inicializace bývá velmi složitá a často je předmětem celých studií [1] a nebudeme se jí věnovat v této práci. Tento výsledek se pak porovná s ukázkovým příkladem pro tento obrázek, tedy pro anotaci, která byla předtím připravena člověkem.

Dále zvolíme metodu pro výpočet chyby výstupu (loss). Proces generování výstupu opakujeme mnohokrát pro velké množství obrázků a anotací toho. Skupině těchto dat říkáme trénovací dataset. Je nutné mít velké množství obrázků a anotací ještě před začátkem trénování. Ty pochází v drtivé většině případů od lidských anotátorů a jsou tak velmi drahé. Navíc jsou předmětem lidské chybovosti. Součtem, nebo jinou metodou tak spočítáme ztrátovou funkci (loss function) jako metriku chybovosti odhadu našeho modelu od anotací

### 1.1.2 Druhy detekcí

K tomuto účelu se využívá hned několik metod. Tou úplně nejzákladnější jsou rámečky k ohraničení detekovaných lidí. skládají se ze dvou souřadnic.

### 1.1.3 Odhad 2D a jeho limitace

Tradiční metody odhadu pózy se po léta soustředily na tzv. sparse keypoints – detekci klíčových bodů, jako jsou lokty, kolena či ramena – které detekují jako dvojici souřadnic v daném obrázku. Přestože jsou tyto modely (např. OpenPose) rychlé a efektivní, trpí zásadním nedostatkem: ztrátou prostorové informace a tělesného objemu. Zatímco mají perfektní výsledky v laboratorních podmínkách, selhávají v reálných situacích (in-the-wild), kde jim 2D obrázek neposkytuje dost jasné informace pro detekci pozice. Mezi faktory ovlivňující přesnost výstupu patří:

- **Zákryty (Occlusions):** Části těla zakryté předměty nebo jinými částmi těla.
- **Perspektivní zkreslení:** Extrémní úhly kamery, které deformují vizuální proporce těla.
- **Vizuální šum:** Volné oblečení nebo špatné světelné podmínky.
- **Propletení více těl:** Situace v davu, kde není jasné, komu patří jaká část těla a jak je správně rozdělit.

### 1.1.4 Od 2D bodů k 3D objemu

Řešení těchto složitých případů významně napomáhá představa 3D modelu těla a jeho pozice na dané scéně. Vezmeme-li situaci člověka skákajícího na lyžích zespodu, jeho tělesné proporce budou zcela nestandardní (malá hlava, velké nohy, spousta zakrytých částí těla). Pokud si ale představím model lidského těla a promítnu ho do obrázku, pak jsem značně omezen a najít správné orientace končetin se značně zjednoduší. Představa těla ve 2D mi naopak dovoluje zvažovat pozice, jež by byli lidské tělo zcela nepřirozené, či anatomicky nemožné. Modely, které spolu s pozicí kloubů odhadují také 3D orientaci těla se díky tomu stávají lepší v samotném odhadování pozice kloubů.

Samotné rozpoznávání 2D pozice se dostává ke svým limitům také proto, že jejich přesnost se blíží datům na kterých jsou trénovány. [2]. Další k rozpoznání 3D pozice těla je zle

## 1.2 Skinned Multi-Person Linear Model (SMPL)

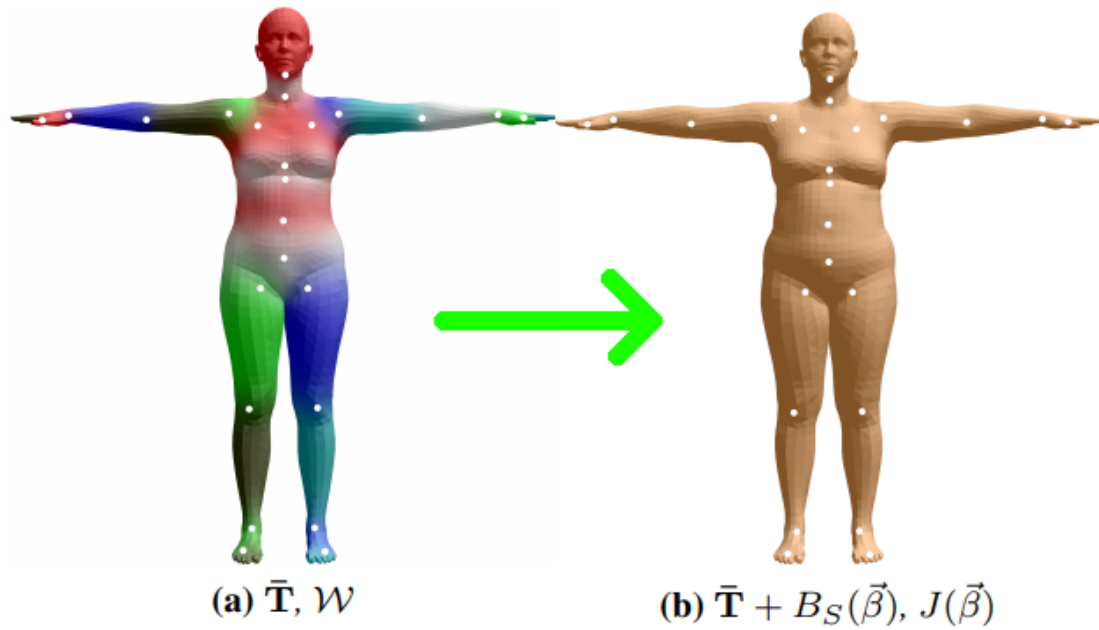
Přechod od 2D chápání člověka k pochopení objemové struktury lidského těla si vyžaduje zcela nové nástroje. Jedním z nich je způsob jak modelovat lidské tělo v prostoru. Abychom mohli tělo vyrendrovat na obrazovku pomocí standardních postupů, musíme povrch těla zapsat jako množinu bodů a zapamatovat si všechny možné trojice bodů tak, aby nám vznikla síťovina (mesh) reprezentující povrch 3D tělesa.

Namodelovat realistické lidské tělo na základě bodů je velmi náročná disciplína. My bychom takové tělo chtěli modelovat automaticky, v reálném čase a s tělesnými proporcemi a pozicí, jakou si zadáme. Proto vznikl SMPL model, který na základě parametrů  $\beta$  a  $\theta$  vytvoří síťovinu libovolného člověka.

### 1.2.1 Výpočet

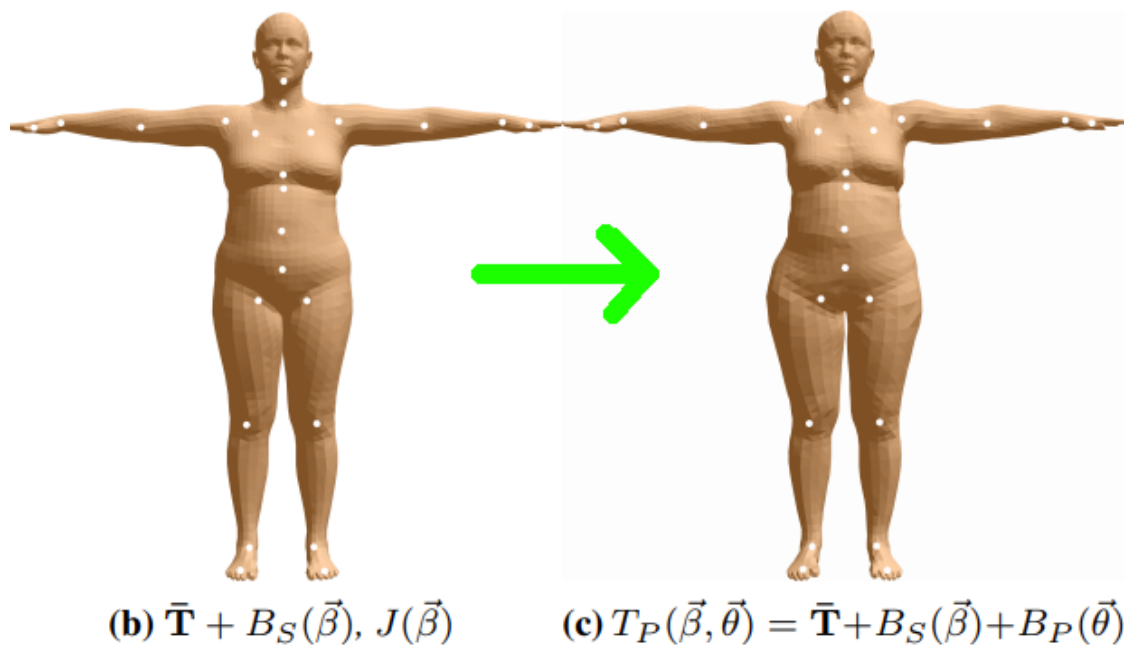
Celý model těla vzniká tak, že se začne se souřadnicemi bodů kanonického těla  $\bar{T}$  v klidu (tedy univerzálního zvoleného těla, které zobrazuje průměrného člověka ve všech proporcích). Ke každému bodu se následně přičte posunutí na základě parametru tělesných proporcí  $\beta$  příspěvkem  $B_s(\beta)$ . Tak vznikne T-pose správných tělesných proporcí. Také vypočteme počáteční pozici kloubů  $J(\beta)$  na základě tělesných proporcí.





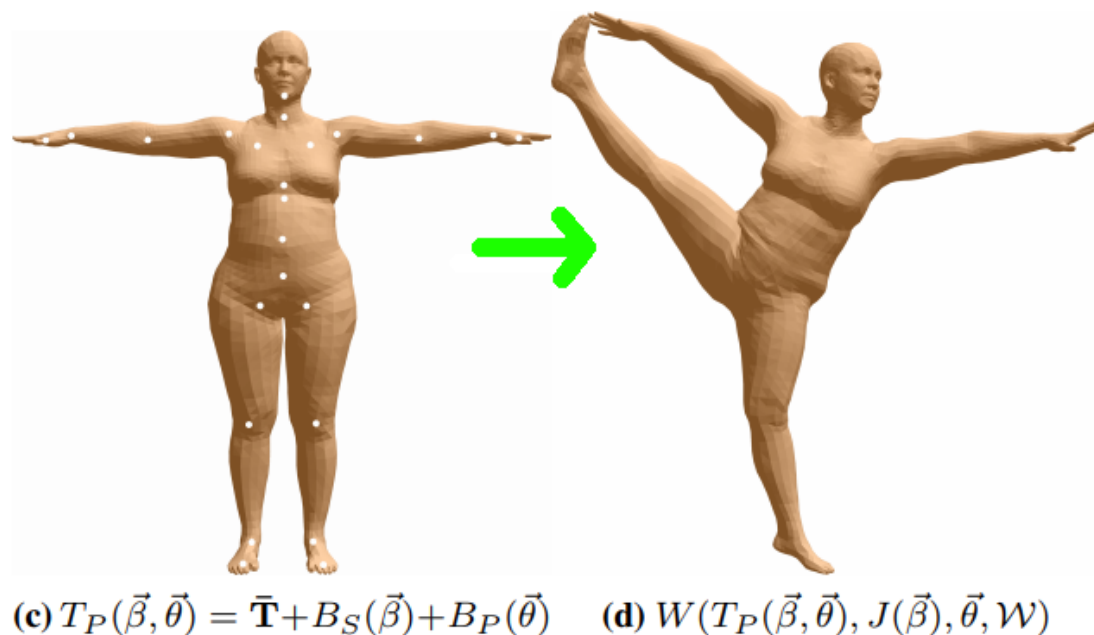
Obrázek 1.1: SMPL – stage 1

Při pohybu se naše tkáně napínají a ohýbají a s tím musíme počítat i u počítačového modelu člověka. Dále tedy upravíme proporce těla tak, aby odpovídali po pozici do které chceme model dostat tím, že k modelu přidáme příspěvek  $B_P(\theta)$  a vznikne tak pozice  $T_P(\beta, \theta)$ .



Obrázek 1.2: SMPL – stage 2

Nakonec se vše poskládá dohromady pomocí funkce  $W(\cdot)$  ještě s konečnou pozicí končetin  $\theta$  a maticí  $\mathcal{W}$ . Tato matice spolu s  $B_S(\beta)$  a  $B_P(\theta)$  jsou natrénované hodnoty na tisících lidských skenů.



Obrázek 1.3: SMPL – stage 3

### 1.2.2 Výhody

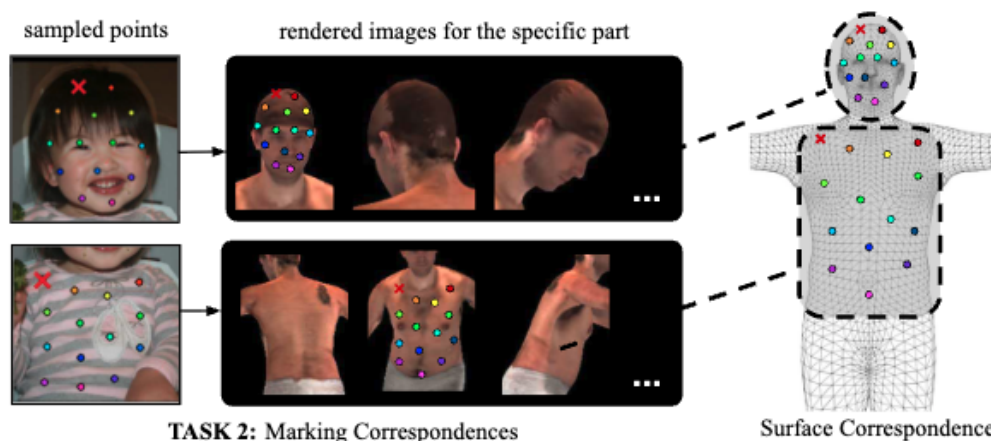
Tento model má výhodu v tom, že parametry  $\theta$  a  $\beta$  usměrní výsledek tak, aby bylo velmi těžké zdeformovat obrázek do nelidských proporcí. Zároveň však dokáže popsat celou řadu lidských těl. Je díky tomu vhodná pro

Tato knihovna je pro komerční účely zpoplatněná, ale pro vědecké účely je zdarma a stačí se registrovat na <https://smpl.is.tue.mpg.de/index.html>. Díky tomu se z ní stal standard v oblasti počítačového vidění.

## 1.3 DensePose

DensePose je metoda z oblasti počítačového vidění, která z obyčejného 2D obrázku odhaduje **hustou korespondenci** mezi pixely a povrchem lidského těla. Na rozdíl od klasického odhadu pózy pomocí klíčových bodů (např. lokty a kolena) se zde nepracuje s několika desítkami bodů, ale s tisíci pixelů, které lze přiřadit ke konkrétním místům na těle. Prakticky to znamená, že pro každý pixel patříčí člověku dokáže DensePose určit, *kam na 3D povrch těla by tento pixel patřil*, pokud bychom měli k dispozici standardizovanou šablonu těla.

Jádrem celé myšlenky je kanonická reprezentace těla (tedy povrch „průměrného“ člověka v jednotných souřadnicích) a síť, která se učí předpovídat mapování z obrazu do této kanonické reprezentace. DensePose je typicky používán jako nadstavba detekce osob: nejprve se nalezne ohraničující rámeček (bounding box) postavy a teprve v jeho rámci se provede husté mapování pixelů na povrch těla. Díky tomu se výpočet soustředí na relevantní část obrazu a je možné pracovat s výstupem po jednotlivých lidech.



Obrázek 1.4: DensePose – DenseAnnoFigure, Zobrazuje korespondenci 2D pixelu s vybraným bodem na kanonické 3D síti lidského těla.

### 1.3.1 DensePose v1: mapování pomocí (I, U, V)

Původní verze DensePose (v této práci ji budu označovat jako DensePose v1) používá parametrizaci povrchu těla pomocí několika tělesných „chartů“ (segmentů). Výstupem je pro každý pixel na postavě trojice (I, U, V):

- **I**: index tělesného segmentu (např. levou paži, trup apod.),
- **U, V**: souřadnice v rámci 2D parametrizace daného segmentu.

Tento výstup je vhodný například pro přenos textur nebo pro vizualizaci mapování povrchu, ale pro následný fitting parametrického modelu (např. SMPL) je potřeba navíc řešit převod z (I, U, V) do konkrétních bodů nebo vrcholů (vertices) na síťovině. To je možné, ale v praxi to přidává další vrstvu komplikací, protože optimalizační algoritmus pak nepracuje přímo s jednoznačnou korespondencí na SMPL mřížce.

### 1.3.2 DensePose v2: Continuous Surface Embeddings (CSE)

Novější varianta, kterou v této práci používám (DensePose v2), přechází od explicitních (I, U, V) souřadnic k tzv. Continuous Surface Embeddings (CSE). Místo toho, aby síť přímo vracela parametrické souřadnice na těle, vrací pro každý pixel vektor embeddingu (tedy bod v naučeném vektorovém prostoru), který reprezentuje odpovídající místo na povrchu těla. Současně vrací i hrubou segmentaci (coarse segmentation), která říká, které pixely v daném rámečku vůbec patří postavě.

Z praktického hlediska tak DensePose v2 produkuje dvě hlavní struktury:

- **Masku S** o rozměrech  $H \times W$  (binární nebo vícetřídní), která určuje pixely patřící tělu.
- **Matici vektorů E** o rozměrech  $H \times W \times D$ , kde  $D$  je dimenze embeddingu, a každý pixel tak nese svůj vektor  $\mathbf{e} \in \mathbb{R}^D$ .

Tyto mapy jsou definované v souřadnicích rámečku detekované osoby a lze je přepočítat na libo-

volné rozlišení (typicky interpolací na nové  $H$  a  $W$ ). V praxi je to důležité, protože výstup se dá škálovat podle toho, zda preferuji rychlost (nižší rozlišení) nebo přesnost a počet korespondenčních bodů (vyšší rozlišení). Pro účely optimalizace je pak možné z embeddingů odvodit i jednoznačný index vrcholu na SMPL síťovině: pro každý pixel se vybere ten vrchol, jehož předpočítaný embedding je vektorově nejbližší (typicky podle kosinové podobnosti). Výsledkem je tedy mřížka indexů, která mapuje pixely přímo na vrcholy SMPL (např. 6890 vrcholů), a tu lze přímo použít v loss funkci.

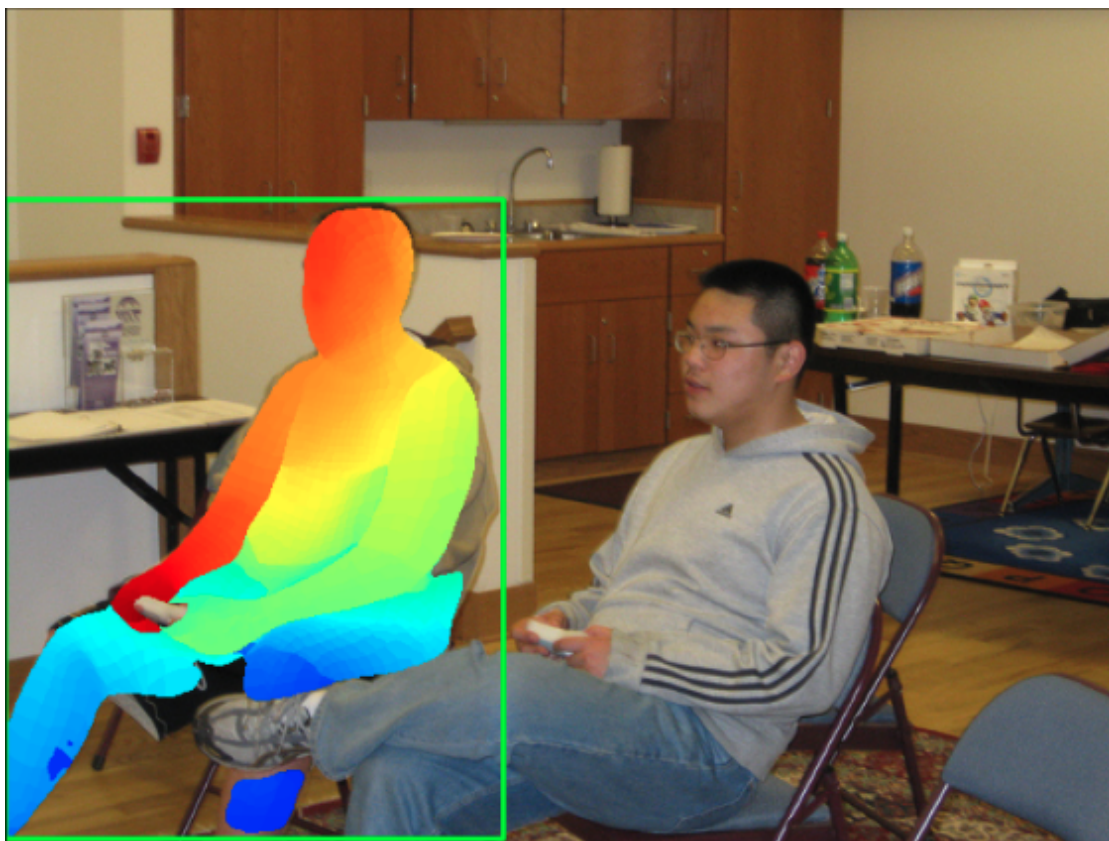
DensePose v2 jsem zvolil z několika důvodů. Především mě překvapila jeho přesnost i v podmínkách mimo laboratorní datasety a zároveň jde o open-source řešení postavené nad ekosystémem Detectron2, které je dobře integrovatelné do vlastního pipeline. Zároveň je férové říct, že DensePose jako projekt dnes působí spíše jako *vědci trochu zapomenutá* technologie: nové práce často preferují holistické modely, které rovnou odhadují parametry těla nebo celé 3D reprezentace. DensePose však zůstává velmi vhodný pro navazující práci právě kvůli tomu, jaký typ výstupu poskytuje. Moderní holistické modely mívají výstup „hotový“ (např. klouby nebo SMPL parametry), ale neposkytují jednoduchý způsob, jak si vyžádat hustou, škálovatelnou korespondenci pixelů na povrch těla, kterou lze přímo zapojit do vlastní optimalizace.

### 1.3.3 Silné a slabé stránky

Největší výhodou DensePose je, že poskytuje hustý signál: i když část těla není jasně vidět, zbytek povrchu často dává dostatek informací pro smysluplné omezení 3D řešení. Tento typ výstupu je navíc přirozeně kompatibilní s fittingem SMPL, protože umožňuje formulovat ztrátovou funkci nad velkým množstvím korespondenčních bodů.

Zároveň má DensePose několik slabin, které se v praxi projeví velmi rychle. Výstup bývá často šumový a místo hladkého povrchu připomíná „lupínky“ (lokálně nekonzistentní přiřazení sousedních pixelů). To vede k tomu, že i při vizuálně správné detekci může být lokální korespondence nekvalitní. Dalším typickým problémem je záměna symetrických částí těla: model může například označit obě nohy jako levé, případně prohodit levou a pravou stranu. V neposlední řadě se chyby objevují i na hranách segmentace (např. u volného oblečení), kde maska zahrne pixely, které ve skutečnosti neodpovídají povrchu těla.

Právě tyto slabiny jsou hlavním důvodem, proč jsem se rozhodl na DensePose navázat. Místo toho, abych DensePose bral jako „konečný“ výsledek, беру jej jako velmi bohatý, ale nedokonalý signál, který je potřeba dále zpracovat a zpřesnit tak, aby byl fitting SMPL stabilní a robustní i v hraničních případech.



Obrázek 1.5: DensePose – Output, Použili jsme barevnou mapu (colormap) Jet pro ilustraci. Ukazuje neplynulou segmentaci. Také artefakt na pravé noze člověka. Obrázek pochází z datasetu COCO.



## 2. Explorativní část

V rámci své explorativní analýzy jsem navrhl 3 možná metodická zlepšení. Všechna z nich využívají metodu napasování (fitting) SMPL postavy.

### 2.1 Vzdálenost bodu na těle

#### 2.1.1 Návrh metody

Tato metoda vychází z myšlenky, že DensePose v2 neposkytuje jen masku postavy, ale především „otisk“ povrchu těla v podobě embeddingu. Každému pixelu, který patří postavě, odpovídá vektor  $\mathbf{e} \in \mathbb{R}^D$ , jenž reprezentuje konkrétní místo na kanonickém povrchu těla. Pokud tedy dokážu pro aktuální odhad SMPL určit, „jaké místo na povrchu“ je vidět v každém pixelu, mohu DensePose použít jako husté omezení (constraint) pro fitting a netrénovat žádnou novou síť.

Klíčovým krokem je získání korespondence mezi pixely v obraze a body na SMPL síťovině. Místo prosté projekce vrcholů do 2D (která ignoruje zákryty a často vede k falešným přiřazením) jsem uvažoval o vyrenderování SMPL pomocí ray tracingu (prakticky ekvivalent Z-bufferu, jen formulovaný jako průnik paprsku s trojúhelníky). Pro každý pixel v rámci bounding boxu tak získám trojici vrcholů trojúhelníku, který je v daném směru nejbližší kameře, a barycentrické váhy průniku. Díky tomu lze pro daný pixel vypočítat „spojitý“ embedding bodu na SMPL povrchu jako interpolaci embeddingů vrcholů.

Formálně označme  $S$  množinu pixelů, které DensePose považuje za tělo (coarse segmentation). DensePose zároveň dává embedding  $\mathbf{e}_p$  pro každý  $p \in S$ . Pro SMPL si předpočítám embedding pro každý vrchol (stejně jako v pipeline DensePose CSE) a při ray tracingu získám pro pixel  $p$  trojúhelník  $(i,j,k)$  a váhy  $(\alpha, \beta, \gamma)$ . Interpolovaný embedding na povrchu SMPL je pak

$$\tilde{\mathbf{e}}_p = \alpha \mathbf{v}_i + \beta \mathbf{v}_j + \gamma \mathbf{v}_k,$$

kde  $\mathbf{v}_i$  jsou embeddingy vrcholů SMPL. Protože embeddingy DensePose se běžně L2-normalizují, dává smysl měřit shodu pomocí kosinové podobnosti, tedy skalárního součinu normalizovaných vektorů (ne vektorového součinu, který je navíc definován jen v  $\mathbb{R}^3$ ). Ztrátovou funkci pak lze napsat například jako

$$\mathcal{L}_{\text{CSE}} = \sum_{p \in S} (1 - \langle \hat{\mathbf{e}}_p, \tilde{\mathbf{e}}_p \rangle),$$

kde  $\hat{\cdot}$  značí normalizaci na jednotkovou délku. Optimalizují se standardní parametry SMPL (globální rotace a posun, parametry pózy  $\theta$  a proporcí  $\beta$ ) tak, aby vyrenderovaný povrch měl v každém pixelu co nejpodobnější „bod na těle“ jako DensePose.

Očekávaný přínos této metody je dvojitý. Za prvé, SMPL funguje jako humanoidní omezení: i když DensePose lokálně šumí nebo vytváří nespojitosti, fitting je nucen najít globálně konzistentní lidskou geometrii a proporce. Za druhé, tímto přístupem lze z jednoho 2D snímku získat nejen lepší odhad kloubů (přes SMPL skeleton), ale i explicitní 3D rekonstrukci těla. V ideálním případě by se tím zároveň potlačila typická selhání DensePose, jako je záměna levé a pravé strany (např. „obě nohy jako levé“), protože takové řešení je pro jedinou SMPL instanci geometricky obtížně realizovatelné.

Nevýhodou je výpočetní a implementační složitost. Aby bylo možné tuto loss funkci skutečně použít v gradientní optimalizaci, je potřeba mít alespoň přibližně diferencovatelný render (diferencovatelný ray tracing / rasterizaci), případně zvolit alternativu, která se vyhne explicitnímu průseku paprsků. Metodu proto беру primárně jako konceptuálně čistý cíl: přímo porovnat „kde na těle“ DensePose říká, že pixel leží, s tím, „kde na těle“ leží daný pixel podle aktuálního SMPL odhadu.

### 2.1.2 Implementace

### 2.1.3 Výsledky

## 2.2 Euklidovská vzdálenost

### 2.2.1 Návrh metody

### 2.2.2 Implementace

### 2.2.3 Výsledky

Tato metoda vychází z toho, že DensePose v2 (CSE) mi pro každý pixel postavy nevrátí jen informaci „tady je člověk“, ale také odhad „kde na těle tento pixel leží“. Konkrétně: DensePose pracuje v rámci ohraničujícího rámečku (bounding boxu) detekované osoby a vytvoří (1) masku postavy a (2) embedding. Masku budu značit jako  $S$ : je to množina pixelů, které DensePose označí jako tělo. Pro každý pixel  $p \in S$  DensePose predikuje vektor  $\mathbf{e}_p \in \mathbb{R}^D$  (embedding), kde  $D$  je pevně daná dimenze (typicky desítky čísel). Intuice je, že podobné vektory odpovídají podobným místům na povrchu těla. Abych tento signál mohl použít pro fitting SMPL, potřebuji umět pro aktuální odhad SMPL určit, jaký „bod na těle“ se v každém pixelu skutečně nachází. To není totéž jako vzít nejbližší promítnutý vrchol, protože projekce do 2D ignoruje zákryty: zadní část těla může být promítnuta do stejné oblasti obrazu jako přední část, ale ve skutečnosti ji kamera nevidí. Proto jsem uvažoval o renderování SMPL z pohledu kamery, ideálně tak, aby pro každý pixel bylo jasné, který trojúhelník sítě je opravdu viditelný (tj. nejbližší kameře). Tohle lze udělat standardně pomocí Z-bufferu, nebo si to lze představit jako ray tracing: pro každý pixel vyšlu paprsek do scény a najdu první průsečík se sítí ovinou. Výsledkem renderování pro pixel  $p$  je trojúhelník na SMPL síťovině, do kterého paprsek narazil. Trojúhelník popíšu třemi indexy vrcholů  $(i, j, k)$ . Zároveň dostanu tzv. barycentrické váhy  $(\alpha, \beta, \gamma)$ , které říkají, jak přesně leží bod průsečíku uvnitř trojúhelníku (platí  $\alpha + \beta + \gamma = 1$ ). Díky tomu umím z vrcholových hodnot dopočítat hodnotu přímo v bodě průsečíku jako „vážený průměr“ vrcholů. DensePose CSE má důležitou vlastnost: pro SMPL vrcholy existuje také předpočítaná tabulka embeddingů. Označím ji jako  $\mathbf{v}_i \in \mathbb{R}^D$ , kde  $i$  je index vrcholu SMPL. Pro pixel  $p$  pak spočtu embedding bodu na povrchu SMPL, který je v daném pixelu vidět, jako

$$\tilde{\mathbf{e}}_p = \alpha \mathbf{v}_i + \beta \mathbf{v}_j + \gamma \mathbf{v}_k.$$

Tím mám dvě porovnatelné věci: DensePose embedding  $\mathbf{e}_p$  (z obrázku) a „SMPL embedding“  $\tilde{\mathbf{e}}_p$  (z vyrenderované sítě). Ztrátovou funkci (loss) pak chci definovat tak, aby byla malá, když oba embeddingy odpovídají stejnému místu na těle. V praxi se embeddingy často normalizují na



jednotkovou délku (L2-normalizace), a potom se podobnost měří kosinovou podobností, což je skalární součin:

$$\langle \hat{\mathbf{a}}, \hat{\mathbf{b}} \rangle,$$

kde  $\hat{\mathbf{a}}$  je vektor  $\mathbf{a}$  po normalizaci (má délku 1). Skvělé na tom je, že hodnota je blízko 1, když vektory míří „stejným směrem“ (tedy jsou si podobné), a menší, když si podobné nejsou. Ztrátu tedy mohou napsat například jako

$$\mathcal{L}_{\text{CSE}} = \sum_{p \in S} p (1 - \langle \hat{\mathbf{e}}_p, \hat{\tilde{\mathbf{e}}}_p \rangle).$$

Význam symbolů: -  $S$  je množina pixelů, které DensePose označilo jako tělo. -  $p$  je jeden konkrétní pixel z této množiny. -  $\mathbf{e}_p$  je embedding z DensePose pro pixel  $p$ . -  $\tilde{\mathbf{e}}_p$  je embedding, který jsem získal z aktuálního SMPL odhadu renderováním a interpolací. -  $\langle \cdot, \cdot \rangle$  je skalární součin (po normalizaci odpovídá kosinové podobnosti). - výraz  $(1 - \text{podobnost})$  je malý, když se embeddingy shodují. Optimalizace pak probíhá stejně jako u běžného fittingu SMPL: měním parametry SMPL tak, aby se minimalizovala ztráta. Prakticky jde o globální rotaci a posun těla vůči kameře, parametry pózy  $\theta$  (ohyby kloubů) a parametry tvaru  $\beta$  (tělesné proporce). Cíl je, aby při renderování SMPL vycházelo pro každý pixel „stejně místo na těle“, jaké predikuje DensePose. Hlavní očekávaný přínos je v tom, že SMPL funguje jako humanoidní omezení: DensePose může lokálně „šumět“ (nespojivosti, přeskokování mezi částmi těla, záměna levé a pravé strany), ale SMPL vždy musí zůstat jedna konzistentní lidská síťovina s realistickými proporcemi. Fitting tak může DensePose využít jako hustý signál, ale zároveň chyby „vyhladit“ tím, že nejlepší řešení musí být globálně lidské. Vedlejším přínosem je, že výsledkem není jen sada 2D bodů, ale rovnou 3D model těla, ze kterého lze odvodit klouby i prostorovou pózu. Nevýhodou je náročnost implementace: pokud chci dělat gradientní optimalizaci, potřebuji, aby renderování (a tím i výběr viditelného trojúhelníku a barycentrické váhy) bylo alespoň přibližně diferencovatelné, jinak je optimalizace nestabilní nebo nefunguje. Metodu proto beru jako „ideální“ formulaci problému: porovnat přímo „bod na těle“ z DensePose s „bodem na těle“ z aktuálního SMPL odhadu v každém pixelu.

## 2.3 Zaměření na přesnost

### 2.3.1 Návrh metody

### 2.3.2 Implementace

### 2.3.3 Výsledky



## 3. Technická dokumentace

Poslední kapitola obsahuje informace o tom, jak projekt, který v rámci maturitní práce vznikl, nainstalovat, spustit a používat.

### 3.1 Ukázka sekce

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

#### 3.1.1 A jedné podsekce

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem

ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

### 3.2 A další sekce

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.



# Závěr

Závěr obsahuje shrnutí práce a vyjadřuje se k míře splnění jejího zadání. Dále by se zde mělo objevit sebehodnocení studenta a informace o tom, co nového se naučil a jak vnímal svou práci na projektu.





# Seznam použité literatury

- [MM16] Dmytro Mishkin a Jiří Matas. *All you need is a good init*. 2016. DOI: 10.48550/arXiv.1511.06422. arXiv: 1511.06422 [cs.LG]. URL: <https://arxiv.org/pdf/1511.06422> (cit. 27.02.2026).
- [Mor25] Šárka Morávková. *Title of the document*. 2025. URL: <https://drive.google.com/file/d/17Mhfoe3TpJCVXy5ry61DvdaYFdGLFuhM/view> (cit. 27.02.2026).



# Seznam obrázků

1.1	SMPL – stage 1 . . . . .	7
1.2	SMPL – stage 2 . . . . .	7
1.3	SMPL – stage 3 . . . . .	8
1.4	DensePose – DenseAnnoFigure, Zobrazuje korespondenci 2D pixelu s vybraným bodem na kanonické 3D síti lidského těla. . . . .	9
1.5	DensePose – Output, Použili jsme barevnou mapu (colormap) Jet pro ilustraci. Ukazuje neplynulou segmentaci. Také artefakt na pravé noze člověka. Obrázek pochází z datasetu COCO. . . . .	11

# Seznam tabulek