

Použití počítačového vidění pomocí NVIDIA Jetson Nano

Tadeáš Těhan¹, Mikuláš Voňka², Jan Pavel Šafrata³

¹Gymnázium Ostrava-Zábřeh; tadeastehan@gmail.com

²Gymnázium Kladno; vonka.mikulas@seznam.cz

³Gymnázium Evolution; honza@klan.cz

Ing. Jakub Klinkovský, Ph.D.; KSI FJFI

Abstrakt

V posledních letech zažívá oblast umělé inteligence a strojového učení dramatický nárůst zájmu a inovací. Tento rozmach otevírá nové možnosti v různých oborech, včetně počítačového vidění. V našem projektu jsme se zaměřili na použití zařízení NVIDIA Jetson Nano, které je vybaveno výkonnou grafickou kartou optimalizovanou pro aplikace umělé inteligence a počítačového vidění. Cílem naší práce bylo rozpoznat pomocí Jetson Nano znaky vlajkového semaforu ¹ na základě klíčových bodů na těle. Po rozpoznání jednotlivých písmen jsme byli schopni dekodovat vysílanou zprávu. Z našich experimentů jsme získali cenné poznatky a diskutovali jsme o praktických výhodách a omezeních tohoto zařízení. Výsledky naší práce se nachází na: https://github.com/tadeastehan/flag_semaphore_jetson.

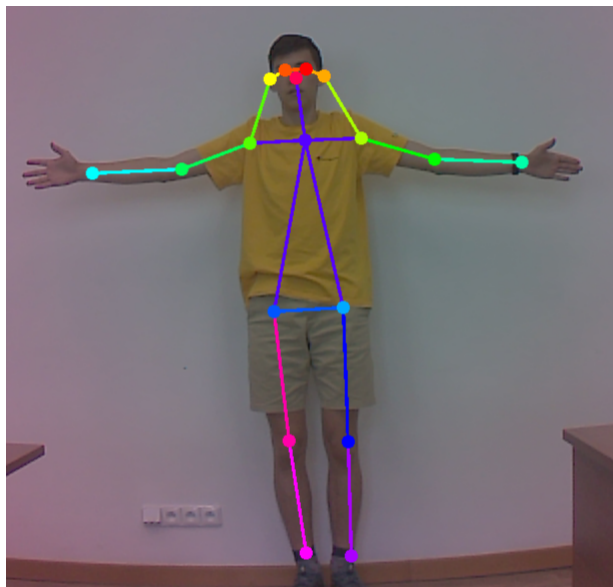
1 Úvod

Zařízení, na kterém jsme celý projekt testovali, byl malý jednodeskový počítač NVIDIA Jetson Nano [2] s GPU². Díky tomu aspektu je náš projekt přenosný a díky akceleraci výpočtů pomocí GPU i dostatečně rychlý. Pro urychlení výpočtů jsme využili technologii CUDA [3] od společnosti NVIDIA, která umožňuje efektivní zpracování paralelních úloh. NVIDIA rovněž poskytuje Software Development Kit [4] pro hluboké učení, což nám značně usnadnilo vývoj aplikace. Na základě našich experimentů jsme analyzovali výsledky a podrobně diskutovali výhody a omezení tohoto zařízení pro praktické použití.

Pro rozpoznávání důležitých bodů na těle jsme použili již předtrénovaný model *Pose-ResNet18-Body*, který dokáže identifikovat 18 klíčových bodů na lidském těle, viz Obrázek 1. Tento model nám umožnil sledovat pozice rukou a efektivně tak rozpoznávat jednotlivé signály semaforu vlajek [1]. Celý systém jsme vytvářeli pomocí programovacího jazyku Python, který nabízí specializovanou knihovnu pro práci s tímto zařízením.

¹Anglicky: *flag semaphore*

²Graphics processing unit



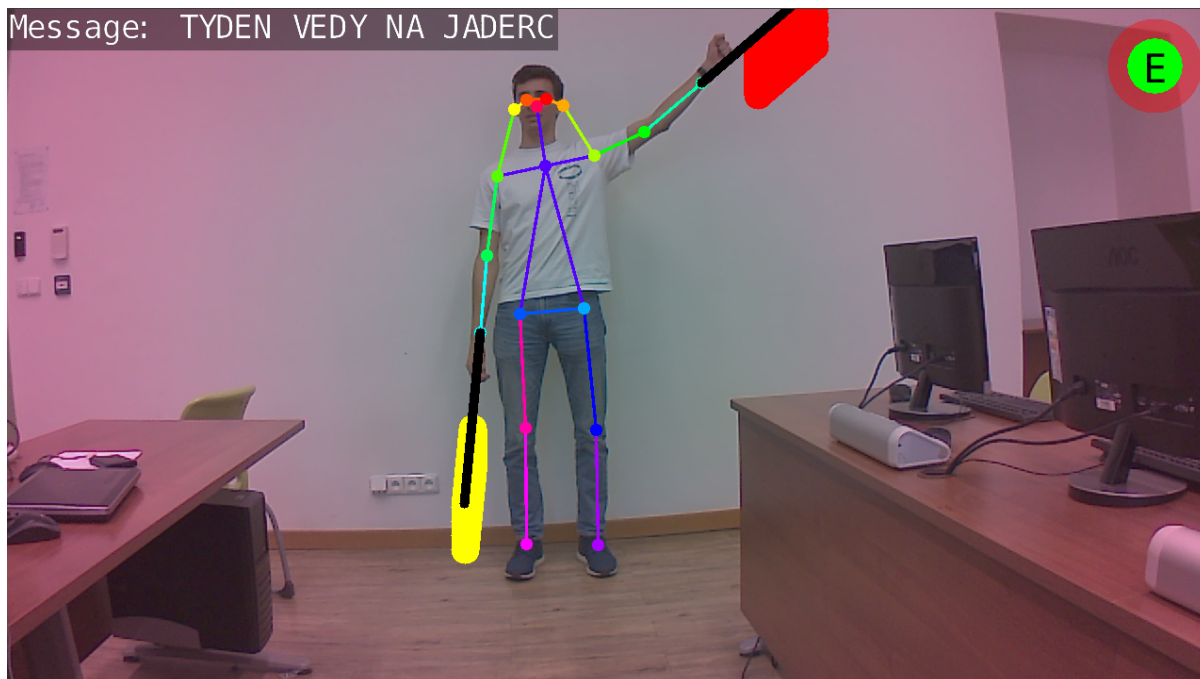
Obrázek 1: Klíčové body na lidském těle detekované neuronovou sítí.

2 Implementace kódu

Naším prvním úkolem bylo umožnit programu rozpoznat gesta. To zahrnovalo identifikaci, kdy jsou paže v rovném postavení, což naznačuje pokus o gesto. Prozkoumali jsme různé metody, jak definovat tuto podmínku, a nakonec jsme našli optimální řešení. Vypočítali jsme rozdíly v souřadnicích x a y mezi body ramene a loktu, stejně jako mezi loktem a zápěstím pro obě paže. Tyto rozdíly nám poskytly vektory různých délek. Když bylo gesto prováděno, tyto vektory byly zarovnané ve stejném směru. Tím, že jsme vektory škálovali tak, aby jejich souřadnice x odpovídaly, jsme zajistili, že pokud by vektory směřovaly stejným směrem, jejich souřadnice y by byly stejné. Poté jsme měřili rozdíl mezi souřadnicemi y dvou vektorů, který by měl být blízko nule, pokud by bylo gesto prováděno.

Následně jsme potřebovali určit úhly, které paže svírají s vertikální osou. Toho jsme dosáhli výpočtem rozdílů v souřadnicích x a y mezi body ramene a zápěstí a následným použitím dvouargumentové funkce atan2 k nalezení úhlu paže vzhledem k vertikální ose. Poté jsme přiřadili hodnoty úhlů pro obě paže ke každému znaku abecedy semaforu vlajek.

Dále jsme vyvinuli podmínku pro přesné detekování písmene, které se osoba snažila signalizovat. Vzhledem k tomu, že spolehnout se na jediný snímek by mohlo vést k chybám, implementovali jsme robustnější přístup. Definovali jsme proměnnou $i = 0$ a referenční proměnnou r jako písmeno z prvního snímku. Dále se v algoritmu mění hodnota r pouze v případě, kdy se i rovná nule. Následná písmena z po sobě jdoucích obrázků byla porovnána s hodnotou r . Pokud se shodovala, zvýšili jsme i o 1; pokud se lišila, snížili jsme i o 1. Když i dosáhne předdefinovaného prahu $2.5 \times f$, kde f představuje počet snímků videa za sekundu, což se stane po 2.5 sekundách signalizace stejného písmene, program potvrdí písmeno a vytiskne ho. Po dosažení tohoto prahu je proměnná i vynulována, referenční proměnná r se změní na aktuální písmeno a začne detekce dalšího písmene zprávy.



Obrázek 2: Grafické prostředí pro dekódování zprávy.

2.1 Grafické prostředí

Součástí grafického prostředí, které je zobrazeno na Obrázku 2, je v levém horním rohu červený kruh, který se postupně v závislosti na proměnné i naplňuje zelenou barvou. Pokud jsme tedy v postavení dostatečně dlouho dobu, tak se kruh naplní zelenou barvou a písmeno se přidá do zprávy v levém horním rohu. Ve středu tohoto kruhu se nám také zobrazuje konkrétní písmeno, které program detekuje. Součástí znaků vlajkového semaforu je i mezera či znak, díky kterému můžeme písmena mazat (backspace). V grafickém prostředí vidíme také kostru rozpoznávaných bodů na lidském těle, které drží virtuální vlajky pro lepší imitaci reality.

3 Výsledky a diskuse

Námi vytvořený program dokáže úspěšně detekovat většinu znaků vlajkového semaforu. Jedno významné omezení nastalo, když byla paže umístěna přímo nad hlavou. Použitá neuronová síť má v této pozici potíže s rozpoznáváním klíčových bodů na paži. Abychom to vyřešili, upravili jsme podmínku detekce gest tak, aby se neaktivovala v určitých úhlech nad hlavou.

Další problém nastal, když paže překřížila tělo a ukazovala na opačnou stranu. Toto umístění někdy způsobilo, že program nesprávně identifikoval klíčové body na paži, což vedlo k chybnému rozpoznání určitých písmen. Implementovali jsme další korekční opatření pro zvýšení přesnosti, ale stále je zde prostor pro zlepšení.

Tato omezení zdůraznila složitost rozpoznávání gest a potřebu neustálého zdokonalování. Budoucí práce by mohla zahrnovat vylepšení algoritmu pro lepší zvládnutí extrémních poloh paží a zkoumání pokročilejších modelů strojového učení. Ke zlepšení přesnosti by mohlo vést natrénování vlastní neuronové sítě přímo pro tyto účely.

4 Závěr

V tomto projektu jsme vyvinuli systém pro rozpoznávání znaků vlajkového semaforu pomocí počítače Jetson Nano. Během vývojového procesu jsme narazili na několik výzev, které vyžadovaly inovativní řešení a úpravy našeho původního návrhu.

Celkově projekt úspěšně demonstroval potenciál využití počítačového vidění a strojového učení pro rozpoznávání gest a znakových jazyků. Navzdory výzvam systém ukázal slibné výsledky a s dalším vylepšením má potenciál stát se spolehlivým nástrojem pro usnadnění neverbální komunikace.

Poděkování

Naše hlavní poděkování si zaslouží Ing. Jakub Klinkovský, Ph.D. za odbornou pomoc při zhotovení práce.

Odkazy

1. WIKIPEDIA CONTRIBUTORS. *Flag semaphore* — *Wikipedia, The Free Encyclopedia*. 2024. Dostupné také z: https://en.wikipedia.org/w/index.php?title=Flag_semaphore&oldid=1228246342. [cit. 2024-06-18].
2. NVIDIA. Jetson Nano. 2024. Dostupné také z: <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/product-development/>. [cit. 2024-06-18].
3. NVIDIA. CUDA Toolkit. 2024. Dostupné také z: <https://developer.nvidia.com/cuda-toolkit>. [cit. 2024-06-18].
4. FRANKLIN, D. Hello AI world guide. 2019. Dostupné také z: <https://github.com/dusty-nv/jetson-inference>. [cit. 2024-06-18].