

Odhad pózy zvířat pomocí metod hlubokého učení

Tomáš Majer¹

1 Úvod

Sledování pohybu divokých zvířat je důležitým aspektem ochrany přírody i lidí. Klasické metody sledování nejčastěji využívají GPS obojky, takový přístup je ale invazivní a umožňuje pozorovat pouze pohyb odchycených jedinců.

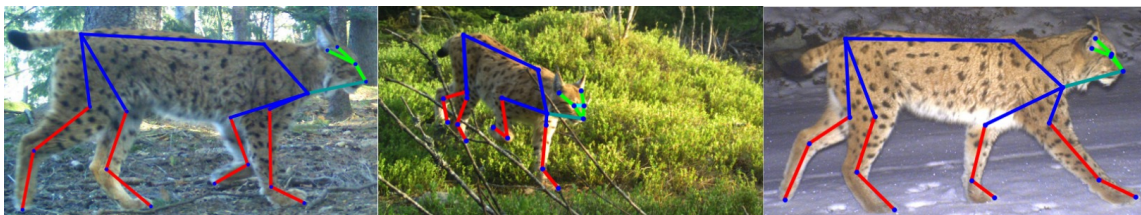
Ke sledování pohybu zvířat se proto postupně začínají používat fotopasti. Využívá se unikátní kresby některých druhů zvířat k jejich identifikaci napříč jednotlivými snímky. Proces identifikace však vyžaduje experta, který jedince rozpozná. Pokud bychom byli schopni kresbu z obrázku získat a vzájemně porovnat s ostatními známými kresbami, proces identifikace by mohl být automatizován. Získání této kresby lze provést například s využitím odhadu 2D pózy zvířat a namapováním textury na tento odhad.

Cílem této práce je návrh funkčního modelu pro odhad pózy zvířat se zaměřením na rysa ostrovida za použití metod hlubokého učení.

2 Estimace pózy rysa ostrovida

V práci byly pro estimaci pózy porovnávány backbone architektury HRNet-W32, ResNet-50 a ResNet-152 s dekonvoluční hlavou produkující heatmapy pro jednotlivé body pózy.

Pracujeme s interním Lynx-Pose datasetem, obsahujícím přes 1600 obrázků rysů ostrovidů anotovaných pro estimaci pózy. Pózu rysa definujeme ve 20 bodech, které jsou ukázány na Obrázku 1 na snímcích z našeho datasetu.



Obrázek 1: Příklad pózy rysů z interního Lynx-Pose datasetu.

V práci jsme nejdříve nastudovali dostupné datasety, které se používají k řešení odhadu 2D pózy zvířat. Jako nejvhodnější dataset byl zvolen dataset Animal-Pose (Cao et al. (2019)), který definuje pózu ve stejných bodech, které používáme v naší úloze.

Jako základní modely jsme použili výše zmíněné sítě implementované v Python toolboxu MMPose a předtrénované na datasetu Animal-Pose. Pro účely experimentů jsme náš dataset rozdělili na trénovací a validační část. V rámci experimentů jsme pak provedli porovnání několika modelů s využitím metriky mean average precision (mAP) měřené na validační části datasetu.

¹ student navazujícího studijního programu Aplikované vědy a informatika, obor Kybernetika a řídicí technika, specializace Umělá inteligence a biokybernetika, e-mail: tomajer@students.zcu.cz

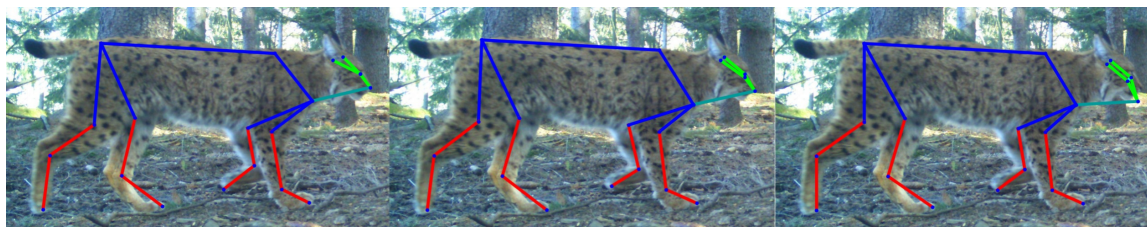
Nejdříve jsme hodnotu mAP změřili pro základní modely předtrénované na Animal-Pose datasetu a následně je natrénovali na Lynx-Pose datasetu. Pozorovali jsme, že kvůli nízkému množství obrázků v trénovací sadě (cca 1200) dochází k přetrénování modelu a to i za použití náhodného zrcadlového překlopení obrázku.

Navrhli jsme tedy pět skupin augmentací (blur, color, noise, weather, geometric) a použili je na trénovací data. Provedli jsme 8 experimentů zkoumajících jejich efekt na úspěšnost modelů. Nejlepší hodnoty mAP dosažené v experimentech jsou v Tabulce 1, kde jsou jako Baseline označeny výsledky základních modelů předtrénovaných na Animal-Pose datasetu. Pro experimenty jsme vždy použili geometrickou augmentaci následovanou další augmentační skupinou.

	Baseline	No augmentations	Blur	Color (C.)	Noise (N.)
HRNet	<u>0.5918</u>	0.8893	<u>0.9042</u>	<u>0.9171</u>	<u>0.9157</u>
ResNet-50	0.4807	0.8802	0.8992	0.9091	0.9079
ResNet-152	0.5330	<u>0.8906</u>	0.9020	0.9134	0.9114
	Weather	C. then N.	N. then C.	C. then N. tuned	C. or N.
HRNet	<u>0.9140</u>	<u>0.9110</u>	0.9122	<u>0.9143</u>	0.9180
ResNet-50	0.9019	0.9080	0.9021	0.9069	0.9137
ResNet-152	0.9089	0.9099	0.9142	0.9119	0.9127

Tabulka 1: Nejlepší hodnoty mAP dosažené jednotlivými modely na validační sadě.

Výsledky estimace pózy nejlepších modelů pro jednotlivé architektury (v Tabulce 1 vyznačeny tučně) jsou pak vidět na Obrázku 2. Jak z vizualizované pózy, tak za použití další metricky (procento správných predikcí na daném vzdálenostním prahu) jsme zjistili, že natrénovaný HRNet model má nejlepší schopnost odhadu pózy.



Obrázek 2: Póza predikovaná pomocí nejlepších modelů (zleva: HRNet-W32, ResNet-50, ResNet-152).

3 Závěr

V rámci práce se podařilo natrénovat model estimace pózy rysa ostrovida. Během práce jsme dospěli k závěru, že architektura HRNet dosahuje v estimaci pózy lepších výsledků než architektury ResNet. Zároveň jsme prozkoumali vliv různých augmentací dat na výsledek trénování.

Literatura

Cao, J., Tang, H., Fang, H. S., Shen, X., Lu, C., Tai, Y. W. (2019). *Cross-domain adaptation for animal pose estimation*. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9498-9507).