

# KNN – Single image depth estimation

František Horázný (xhoraz02)

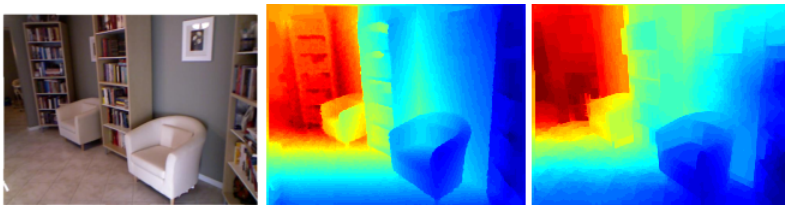
Ondřej Pospíšil (xpospi0a)

Pavel Ševčík (xsevci63)

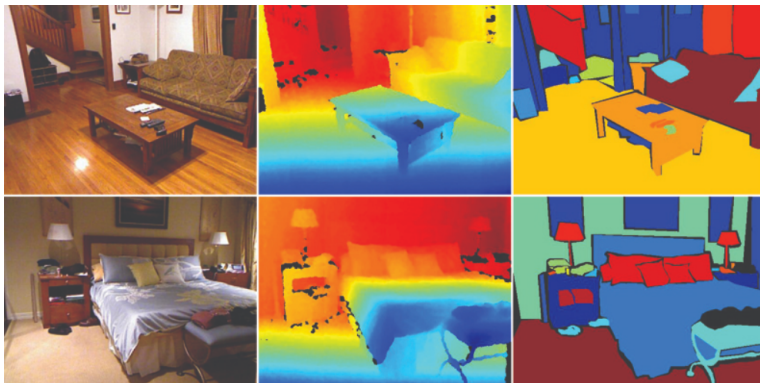
Fakulta informačních technologií Vysokého učení technického v Brně  
Božetěchova 1/2. 602 00 Brno - Královo Pole



- Odhad hloubky každého pixelu snímku
- Datová sada NYU Depth V2
- Metoda AdaBins

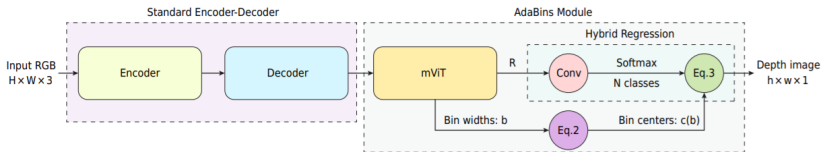


- Snímky místností a vnitřních prostor
- 120 tisíc snímků
- Hloubka do 10 metrů

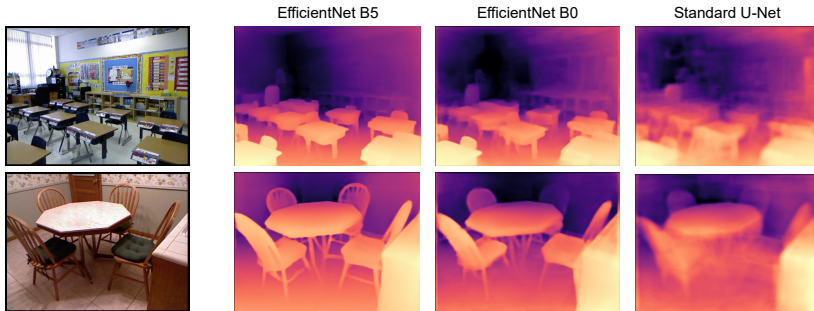


<sup>0</sup>[https://cs.nyu.edu/~silberman/datasets/nyu\\_depth\\_v2.html](https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html)

- Depth Estimation using Adaptive Bins<sup>1</sup>
- Současně nejlepší metoda
- Složena ze dvou základních částí:
  - 1 Encoder-Decoder
  - 2 AdaBins modul



<sup>1</sup>Bhat, Shariq Alhashim, Ibraheem Wonka, Peter. (2020). AdaBins: Depth Estimation using Adaptive Bins.



**Obrázek:** Porovnání výstupů při použití enkodérů EfficientNet B5, EfficientNet B0 (3) a U-Net (2).

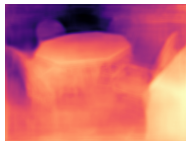
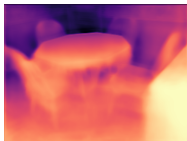
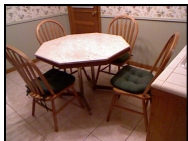
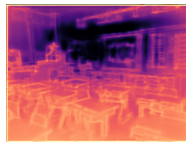
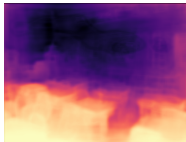
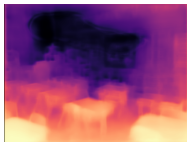
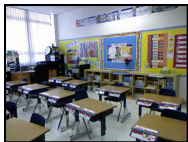
| Enkodér-dekodér     | Podintervaly | Rychlost infer. | Velikost | $\delta_1$ | $\delta_2$ | $\delta_3$ |
|---------------------|--------------|-----------------|----------|------------|------------|------------|
| EfficientNet B5 (3) | 256          | 11.53 sním./s   | 940 Mb   | 0.903      | 0.984      | 0.997      |
| EfficientNet B0 (3) | 256          | 18.20 sním./s   | 326 Mb   | 0.836      | 0.974      | 0.995      |
| U-Net (2)           | 256          | 27.04 sním./s   | 131 Mb   | 0.721      | 0.932      | 0.984      |
| U-Net               | 128          | 34.84 sním./s   | 130 Mb   | 0.728      | 0.936      | 0.983      |

**Tabulka:** Shrnutí výsledků experimentů s architekturou enkodéru-dekodéru a počtem používaných podintervalů. Rychlost je měřena na grafické kartě RTX 2070 Super.

Standard

Depthwise separable

75% velikost vstupů



**Obrázek:** Porovnání výstupů při použití sítě pro extrakci příznaků U-Net (2) a 128 podintervalů. Modely se liší v typu použitých konvolucí a ve velikosti vstupního obrázku.

| Typ konvoluce           | Velikost vstupu | Rychost infer. | Velikost | $\delta_1$ | $\delta_2$ | $\delta_3$ |
|-------------------------|-----------------|----------------|----------|------------|------------|------------|
| obyčejná                | originální      | 34.84 sním./s  | 130 Mb   | 0.728      | 0.936      | 0.983      |
| depthwise separable (1) | originální      | 34.85 sním./s  | 78 Mb    | 0.689      | 0.919      | 0.981      |
| depthwise separable     | 75%             | 52.3 sním./s   | 55.9 Mb  | 0.298      | 0.627      | 0.847      |

**Tabulka:** Shrnutí výsledků experimentů s typem konvolucí a velikosti vstupních obrázků při použití enkodéru-dekodéru U-Net a 128 podintervalů.



- Původní implementace architektury AdaBins je velmi náročná na trénování, má nízkou rychlost inference a výsledný model je velký.
- Povedlo se nám zvýšit rychlost inference, snížit paměťové nároky a zároveň snížit dobu trénování.
- Např. model v nastavení U-Net, redukovanými podintervaly hloubky a depthwise separable konvolucí dosahuje 3x rychlejší inference a 12x menší velikosti. I přesto produkuje kvalitní výsledky o přibližně 1.5 až 21 % nižší v závislosti na metrice v porovnání s původním modelem.

Děkujeme za pozornost.



F. Chollet.

Xception: Deep learning with depthwise separable convolutions.

2017.



O. Ronneberger, P. Fischer, and T. Brox.

U-net: Convolutional networks for biomedical image segmentation.

*CoRR*, abs/1505.04597, 2015.



M. Tan and Q. Le.

EfficientNet: Rethinking model scaling for convolutional neural networks.

In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.