НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Факультет компьютерных наук

Faculty of Computer Science
School of Software Engineering

Bachelor’s Thesis work on the theme:

Алгоритм для создания карт глубины для устройств с ограниченными ресурсами
Novel Algorithm for Depth Map Estimation for Resource-Limited Devices /research work/

Written by the student of the group БПИ171
Bachelor’s education programme
09.03.04 «Software Engineering»
Dagvanorov Lkhagvajav

Supervised by
Ilya Makarov Andreevich, PhD
School of Data Analysis and Artificial Intelligence

Moscow
2021

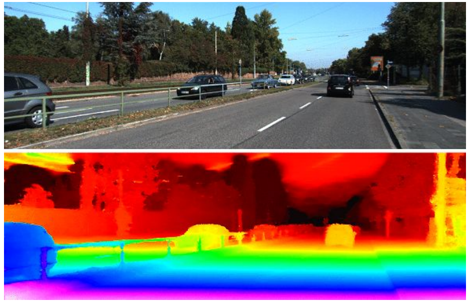
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Content

- Description of the Research Field (1)
- Description of the Research Field (2)
- Motivation (1) (2)
- Goal and tasks
- Methodology: Neural Architecture Search, ProxylessNAS
- ProxylessNAS: Our proposed design / search space
- Experiment settings
- Used frameworks, Libraries and Tools
- Experiment result (search)
- Experiment result (training)
- Result and Comparison / Analysis
- Conclusion and Recommendation for further work
- Reference

Description of the Research Field (1)

- This research project is a subfield of Computer Vision of Machine Learning.
- Depth map is visual data to represent the distance between surrounding objects and the camera/sensor.
- Depth estimation has been applied in many industries as a vital part, since it helps to various ML tasks (object recognition, semantic segmentation, visual construction). To mention: Autonomous driving vehicles(car, robots, drones);



(Top) RGB image taken from a camera
(Bottom) Corresponding depth map

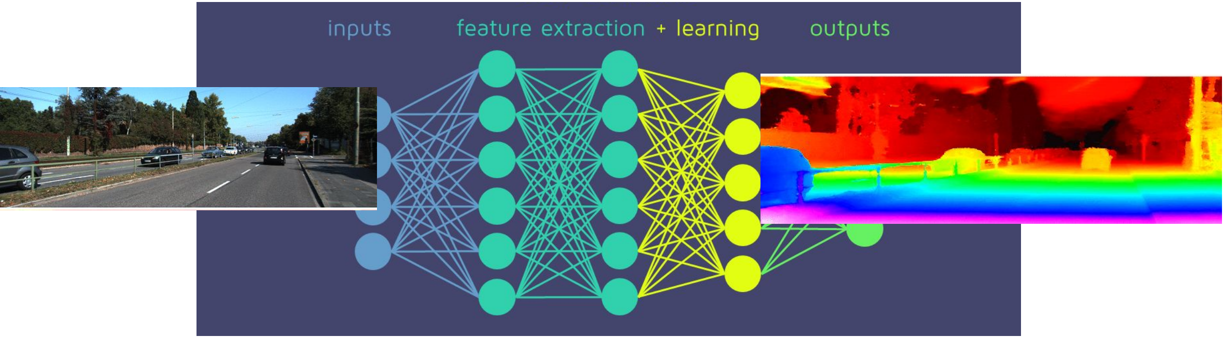
(Left) RGB image taken in classroom (Right) Corresponding depth map.
Redness increases as far distance

Drone picture

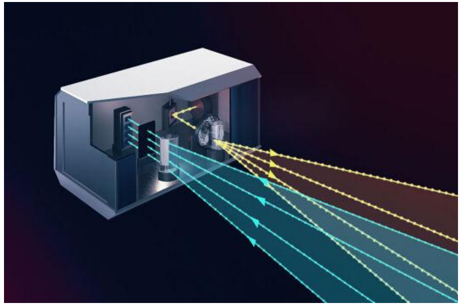
Description of the Research Field (2)

- Today, there are 2 main approaches are used to estimate depth:
  - Acquiring through specific laser sensors (LiDAR, Kinect, etc. 1970-present)
  - Deep Learning Approach (2014-present)

Deep Learning approach. Takes



Deep Learning Approach. Receives RGB image as input and output is the depthmap



Lidar camera. Sends laser beam through yellow vector, and receives the depth information through the blue vector.

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Motivation (1)

- (Motivation-1) Since 2014, the research on deep learning approach of depth estimation has been increased largely as computational resource (GPU, data) have become widely available.

  Despite having great successes in the development, there is one drawback on applying them broadly - “resource-constraints”. Most state-of-art models require high computational resource and not every devices are capable of running them models on itself.

Motivation (2)

- (Motivation-2) In the last 10 years, Machine Learning (ML, DL) based solutions have been used in entirely almost all fields (healthcare, finance, transportation, security, etc.) However using ML requires depth knowledge about the specified field, there is a rising demand for ML specialists.
  Neural Architecture Search is a groundbreaking new idea that it would allow people to create ML models with small or no knowledge about the field. Since 2018, it has been studied and several achievements were made.
  But due to the its complexity to understand and computational requirements it has not been studied widely.

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Purpose and tasks of the research

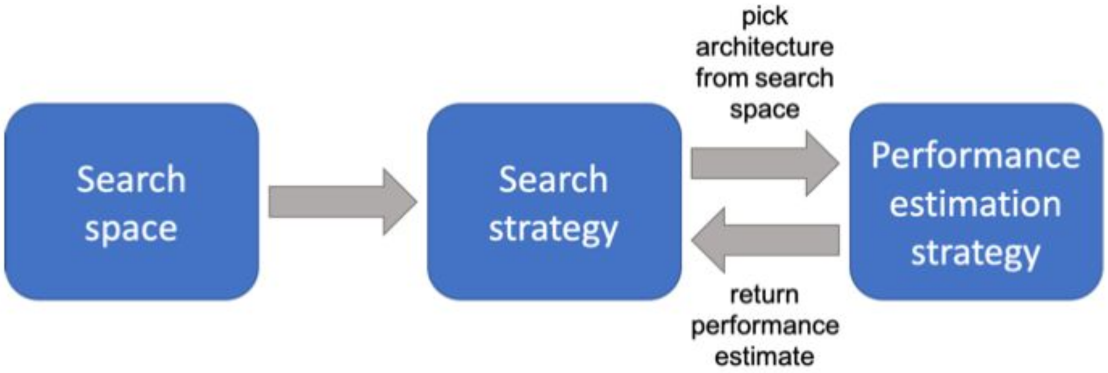
Purpose (goal) of the work: This work has been made as an “comparative research” to investigate new insights of the designing depth estimation model with Neural Architecture Search.

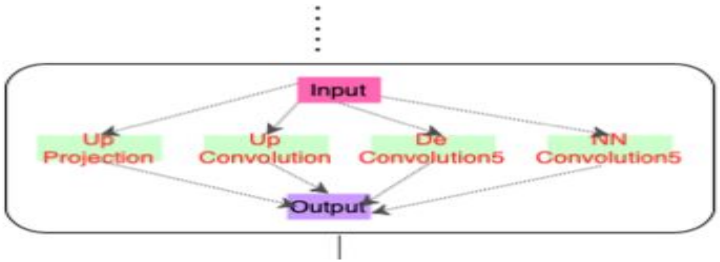
Our initial hypothesis: “Applying NAS on depth estimation will give comparative and promising result.”

Tasks:

1. Exploration of domain and existing approaches (depth estimation, NAS).
   a. Analyze and compare the approaches.
   b. Perform empirical evaluation of the methods that have been developed.
   c. Select ones which are most suitable for our case.
2. Develop/propose a new approach
   a. Propose a new search space
3. Find tools, dataset, and utilities for investigation (experiment)
4. Run experiments
   a. Start search process
   b. Train founded networks
5. Compare and analyze results

Neural Architecture Search, ProxylessNAS



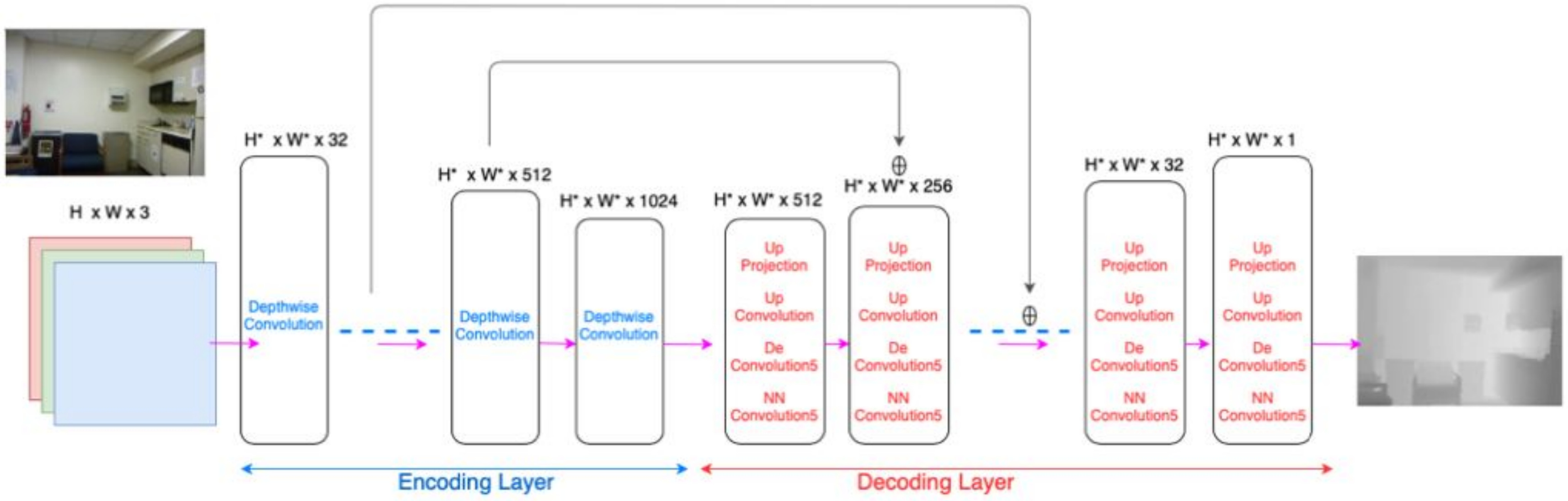


Neural Architecture Search, ProxylessNAS

Our algorithm (approach) can be understood in the following way:

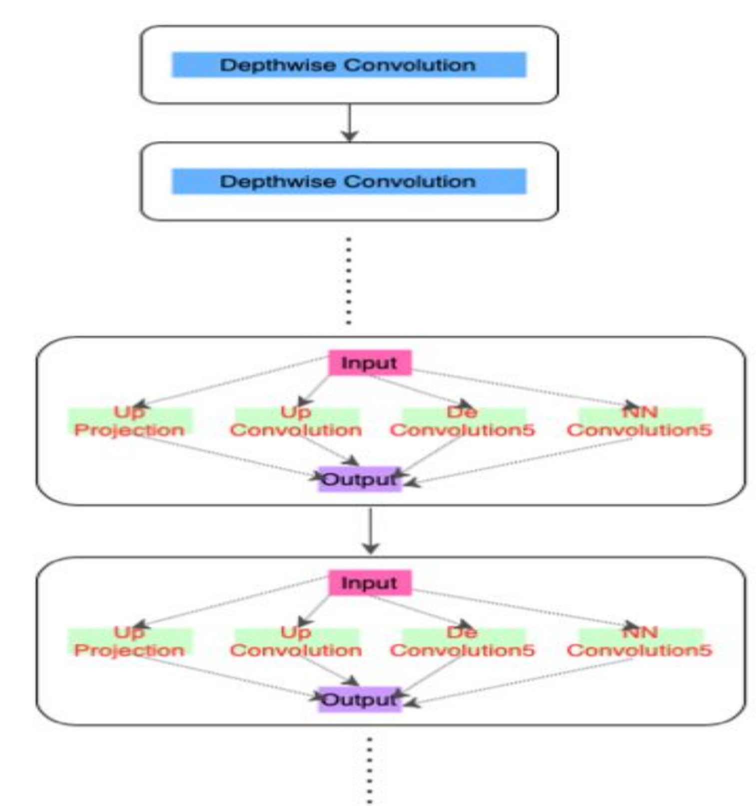
1. **Determine the search space**
   - **Search space is a set of all possible neural operators (convolutional layer, identity layer, etc.)**
2. **Generate a neural network design using ProxylessNAS and the search space**
3. **Train the generated model and evaluate.**

ProxylessNAS: Our proposed design / search space

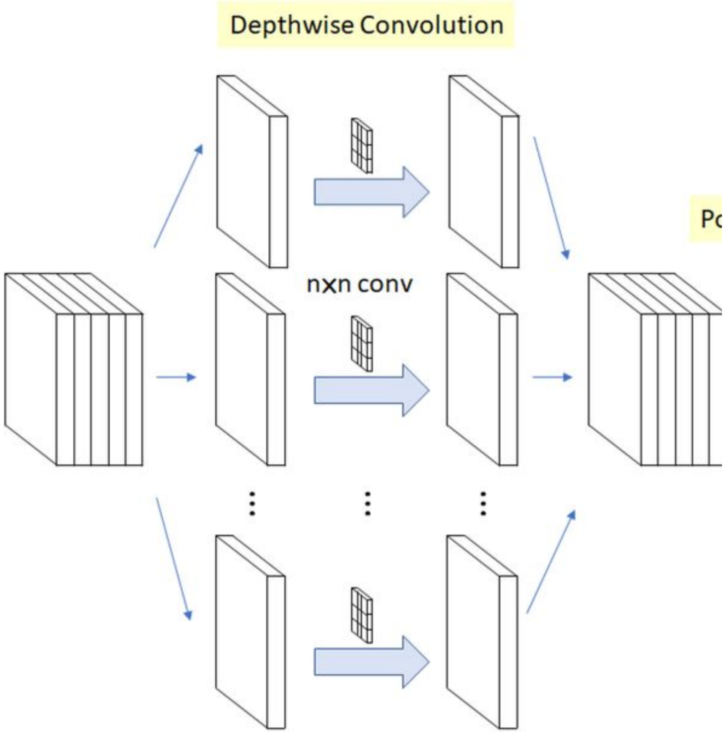


- (Blue Part) The encoder layers are used to extract "depth cues" from the input. Through the layers, the spatial dimension is reduced (480x480x3 -> ~30x30x40)
- (Red Part) The decoder layers construct the spatial reduced image to the original size (heigh, width).

ProxylessNAS: Our proposed design / search space



ProxylessNAS: Our proposed design / search space



Experiments settings

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

- Choosing the dataset:
  - There are currently 2 applicable datasets are available to public: NYUDepth and KITTI. Since searching and training processes take long hours, we’ve decided to choose NYUDepth. The size of the dataset is roughly ~465gb, and we’ve created smaller one by sampling from it.
  - The following transportations are taken on
    - the training set: 1. Resize 2. Rotate 3. Resize 4. Center-Crop 4. Horizontal-Flip 6. Resize
    - the validation set: 1. Resize 2. Center-Crop 3. Resize

- Searching process (ProxylessNAS):
  - Loss function: RMSELoss
  - Optimizer: SGD (learning rate=0.05, momentum=0.1, nesterov=True)
  - Train size = 4800 scenes, Validation = 900, number of epochs = 100

- Training process:
  - Loss function: RMSE
  - Optimizer: Adam (learning rate=0.001, betas=(0.9, 0.999), eps=1e-08, weight_decay=4e-5)
  - Train size = 4800 scenes, Validation = 900, number of epochs = 50

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Used Frameworks, Libraries and Tools

- All experiments were carried on the HSE's High Performance Cluster. And all files of the project (code, script, trained weights of models, etc.) are located at the address: dlkhagvazhav@cluster.hpc.hse.ru

- Microsoft Neural Network Intelligence (NNI) (https://github.com/microsoft/nni) – A completely new framework for applying Neural Architecture Search.

- PyTorch (https://pytorch.org)

- And supplementary packages (h5py, opencv, scikit-image, etc.)

Neural Network Intelligence

PyTorch

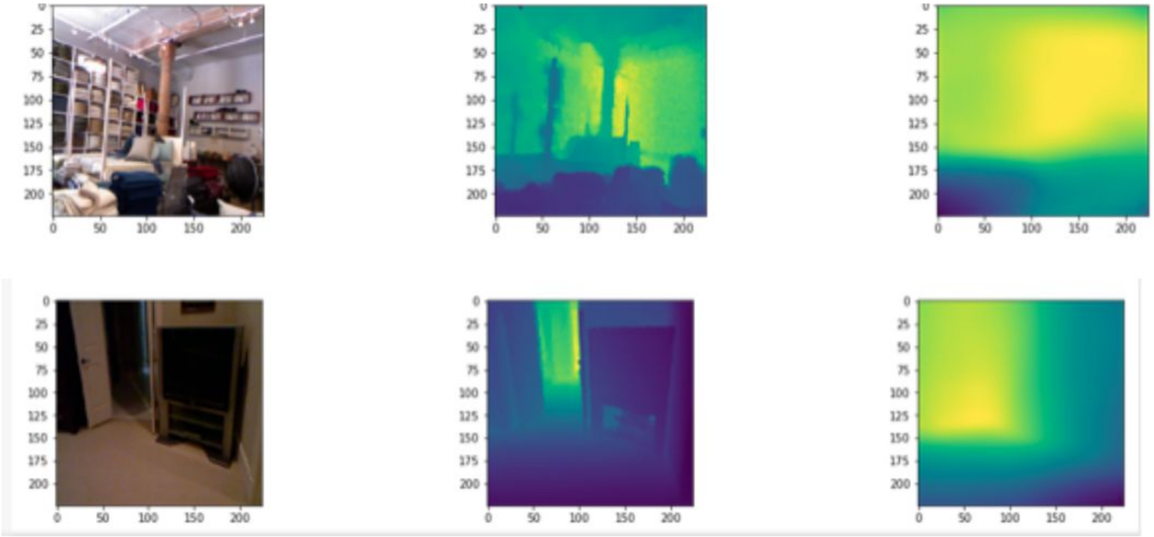
slurm
workload manager

Experiment result (search)

	Candidates	ProxylessNAS choice - Final model for training
Encoder	L1: ConvBn(3, 32, 2) L2: ConvDw(32, 64, 1) L3: ConvDw(64, 128, 2) L4: ConvDw(128, 128, 1) L5: ConvDw(128, 256, 2) L6: ConvDw(256, 256, 1) L7: ConvDw(256, 512, 2) L8: ConvDw(512, 512, 1) L9: ConvDw(512, 512, 1) L10: ConvDw(512, 512, 1) L11: ConvDw(512, 512, 1) L12: ConvDw(512, 512, 1) L13: ConvDw(512, 1024, 2) L14: ConvDw(1024, 1024, 1)	Default same encoders
Decoder	L15: UpProj(1024, 512) / DeConvDw(1024, 512, 5) / NNConvDw(1024, 512, 5) / UpConv(1024, 512)	UpProj(1024, 512)
	L16: UpProj(512, 256) / DeConvDw(512, 256, 5), NNConvDw(512, 256, 5) / UpConv(512, 256)	UpProj(512, 256)
	L17: UpProj(256, 128) / DeConvDw(256, 128, 5) / NNConvDw(256, 128, 5) / UpConv(512, 256)	UpProj(256, 128)
	L18: UpProj(128, 64) / DeConvDw(128, 64, 5) / NNConvDw(128, 64, 5) / Upconv(512, 256)	UpProj(128, 64)
	L19: UpProj(64, 32) / DeConvDw(64, 32, 5) / NNConvDw(64, 32, 5) / UpConv(512, 256)	UpProj(64, 32)
	L20: PointWise(32, 1)	PointWise(32, 1)

Experiment Result (Training)

- Single ready to use model requires roughly 18 hours of search and 12 hours of training processes.
  (Left) RGB image taken from a camera (Middle), Ground truth acquired using Microsoft Kinect (Right) Depth maps calculated from our experiment



Result and Comparison / Analysis

- From the table below, we can see the comparison with our best model and several state-of-art models in several metrics, including RMSE/Delta1 for accuracy and most importantly Latency.
- Our best model (out of 3) has shown promising result in Latency, despite the accuracy was not accurate as other models.

Model	RMSE (lower better)	Delta1 (higher better)	Latency of processing per image (ms)
Eigen[1]	~0.907	~0.611	23
Zhou[2]	~1.04	~0.305	
Xian [5]	~0.66	~0.811	283
Our Best	**~1.0**	**~0.43**	**4**

Conclusion and Recommendation for further work

Although it was ambitious it was also risky attempt to using a completely new approach, Neural Architecture Search to solve the problem.
To best our knowledge, there is no work has been made on applying Neural Search Algorithms on Depth Estimation task. Thus, it is the very first attempt.
Also, there are not many NAS application works were made due to the computational requirements.

Conclusion and Recommendation for further work

From our experiments, it can be advised that use “numerous layers” as possible.

And the conclusion is that we have used NAS approach to design an architecture of depth estimation neural network and it has shown promising results. I succeeded in the goal in the sense of convenient neural network for resource-limited devices.

Thus, I can advice that applying this new approach, NAS algorithm is beneficial if we consider the computational resource required for the search process.

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Reference

[1] D. Eigen, C. Puhrsch, R. Fergus, Depth map prediction from a single image using a multi- scale deep network, in: Advances in Neural Information Processing Systems 27, 2016, pp. 2366–2374.

[2] W. Zhuo, M. Salzmann, X. He, M. Liu, Indoor scene structure analysis for single image depth estimation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 614–622.

[3] H. Cai. ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware. 2019

[4] D. Work. FastDepth: Fast Monocular Depth Estimation on Embedded Systeme. MIT. 2019

[5] K. Xian, C. Shen, Z. Cao, H. Lu, Y. Xiao, R. Li, and Z. Luo, “Monocular relative depth perception with web stereo data supervision,” in Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 311– 320.

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Факультет компьютерных наук

Thank you so much for your attention

dlkhagvazhav@edu.hse.ru

Question Answering Minutes