Depth selectivity in MDE

A study on METER architecture

Master’s Degree in Artificial Intelligence and Robotics

Odysseas Diamadis (1762553)

Academic Year 2024/2025

Sapienza Università di Roma

Outline

o

- Definition of selectivity in MDE
- Describe a technique to enhance selectivity in any model
- Define the experimental setup
- Conclusions and future works

Neuron Selectivity

o

Definition: Neuron selectivity is the ability of a hidden unit of a DNN to react to specialized stimuli, or inputs.

By analyzing neural selectivity it is possible to gather more accurate insights on a network.

Selectivity in MDE

Other works

- **THIRDEYE** Inspired by biological neuron mechanisms, it imitates neuron selectivity by employing layers that that exploit the contours and borders of objects and uses memory mechanisms similar to neural activity during working memory.
- **DSCNet** employs neuron selectivity by allowing the network to exploit semantic information from feature maps, and then integrating them through a module called Multi-level Feature Aggregation Module.

These methods are built on specialized architecture.
Is it possible to build a more generalized method?

The method

Quantifying the average depth response

Definition: $R^d_{l,k}$ as the average response of unit k in layer l on depth d

$$R^d_{l,k} = \frac{\sum_{i=1}^{N} S\left(\tilde{A}_{l,k}(x_i) \odot M_i^d\right)}{\sum_{i=1}^{N} S\left(M_i^d\right)} \tag{1}$$

- $\tilde{A}_{l,k}(x_i)$: upsampled activation map of unit k on input image x_i
- M_i^d: binary mask indicating pixels with discretized depth value d in image i
- $S(\cdot)$: sum over all elements of a matrix
- $\odot$: element-wise multiplication

The method
Training with selectivity in mind

IDEA:

- Divide the depth space in discrete bins
- Assign a specific bin to each hidden unit

$$d_{l,k} = \left\lfloor \frac{k}{K_l / N_b} \right\rfloor \tag{2}$$

l index of current layer, k index of the hidden unit in layer l, Kl the number of hidden unit of layer l, N_b number of bins.

The method

Training with selectivity in mind

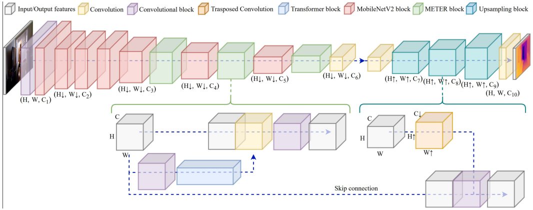
The assignments allow us to define the following objective function

$$\mathcal{L}_{assign} = -\lambda \sum_{l \in L} \frac{1}{K_l} \sum_{k} \frac{|R_{l,k}^{d_k}| - |\bar{R}_{l,k}^{-d_k}|}{|R_{l,k}^{d_k}| + |\bar{R}_{l,k}^{-d_k}|} \tag{3}$$

Where

- $|R_{l,k}^{d_k}|$: Average response on the assigned depth bin
- $|\bar{R}_{l,k}^{-d_k}|$: Average response on all the other bins
- λ is a hyperparameter (0.1 in the original paper)

METER: a mobile vision transformer architectture for MDE



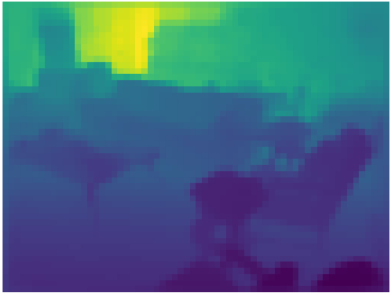
Training procedure reproduced according to the original METER paper

- **xxs** variant
- **Batch size:** 32
- **Optimizer:** AdamW
  - Initial learning rate: 0.001
  - Learning rate schedule: 0.1 reduction every 20 epochs

NYU Dataset

o

- A set of 50k indoor scenes their respective depth maps
- Depths up to 10 meters.



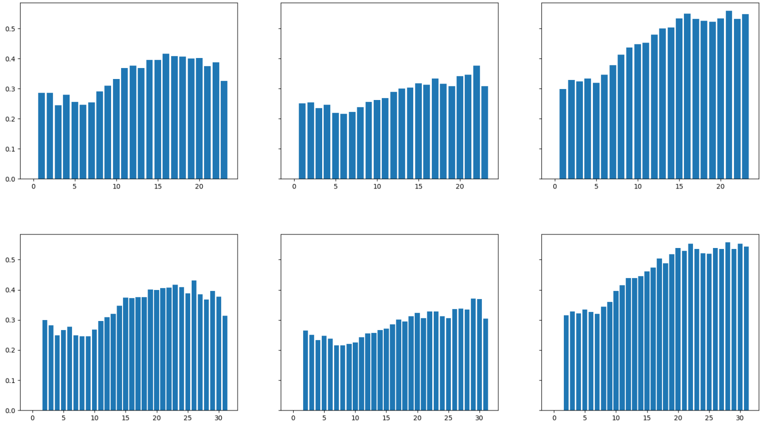
Results

o

	RMSE	REL	$\delta_{1.25}$	λ	N_d
exp1	**64.251**	**0.185**	**0.713**	–	–
exp15	68.491	0.199	0.678	10	32
exp16	69.231	0.200	0.676	100	32
exp17	67.766	0.197	0.683	10	24
exp18	68.214	0.196	0.687	100	24

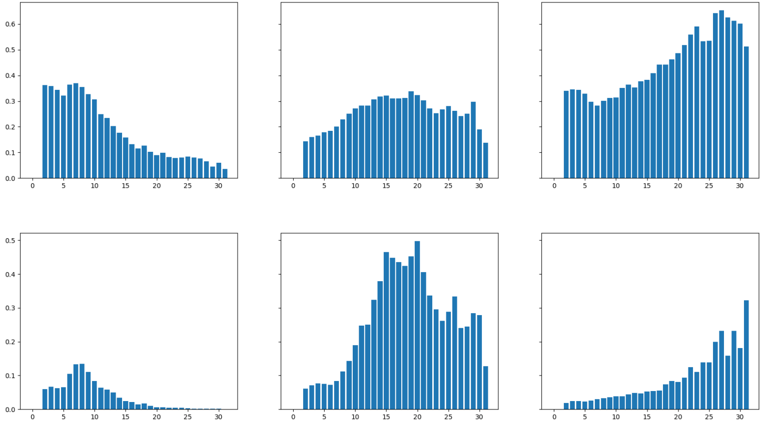
Average response on layer 6

No selectivity, units 12, 24, 40



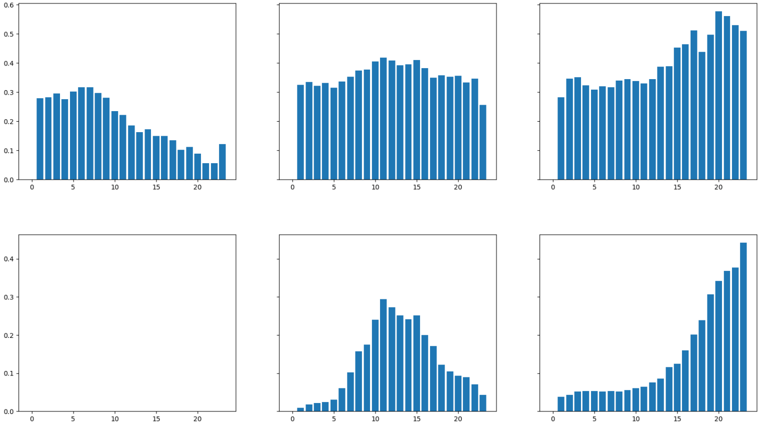
Average response on layer 6, $N_d = 32$

$\lambda = 10, 100$ respectively, units 12, 24, 40



Average response on layer 6, $N_d = 24$

$\lambda = 10, 100$ respectively, units 12, 24, 40



Conclusions

o

- Training based on selectivity negatively impacts performance.
- The value of λ should be tuned with respect to the output magnitude of the model.
- The number of discretization bins significantly affects the depth selectivity of units.
- **Future work:** explore different strategies to optimize the depth assignment formula d_k, potentially by selecting a different discretization scheme on each layer.

Depth selectivity in MDE

Thank you for listening!