Tunnel Effect in CNNs: Image Reconstruction From Max-Switch Location

March, 5th 2018 v0.1

Matthieu de La Roche Saint-André
Akeneo, SIIT

You Deserve Ugliness

You deserve ugliness

Instant Ugly Face

Dataset of caricatures

Our dream:

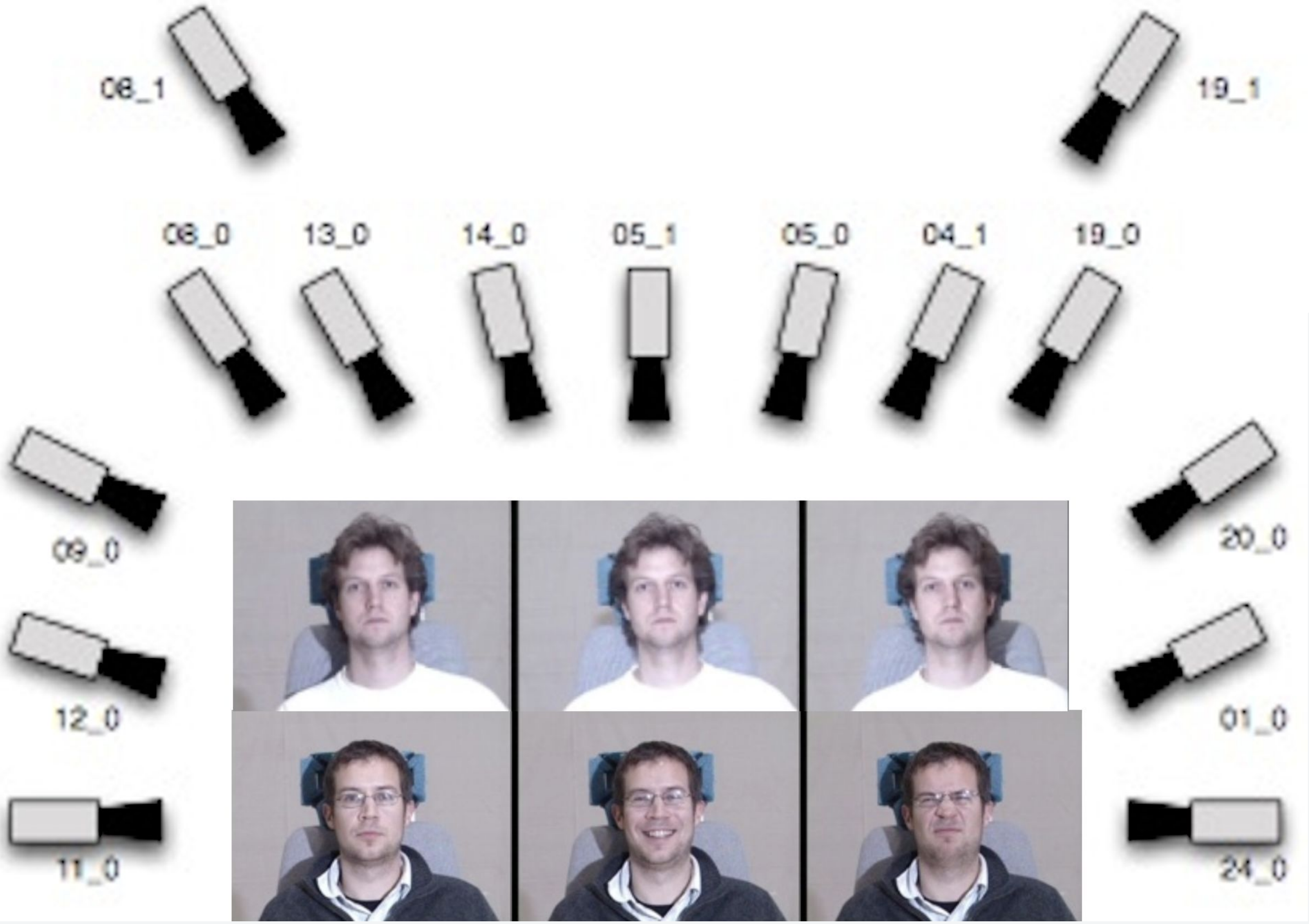
Reality back in 2015: **Nothing**

=> Unsupervised Learning

Since 2017: **WebCaricature**
Huo, Jing, et al.
"WebCaricature: a benchmark for caricature face recognition."
https://cs.nju.edu.cn/rl/WebCaricature.htm

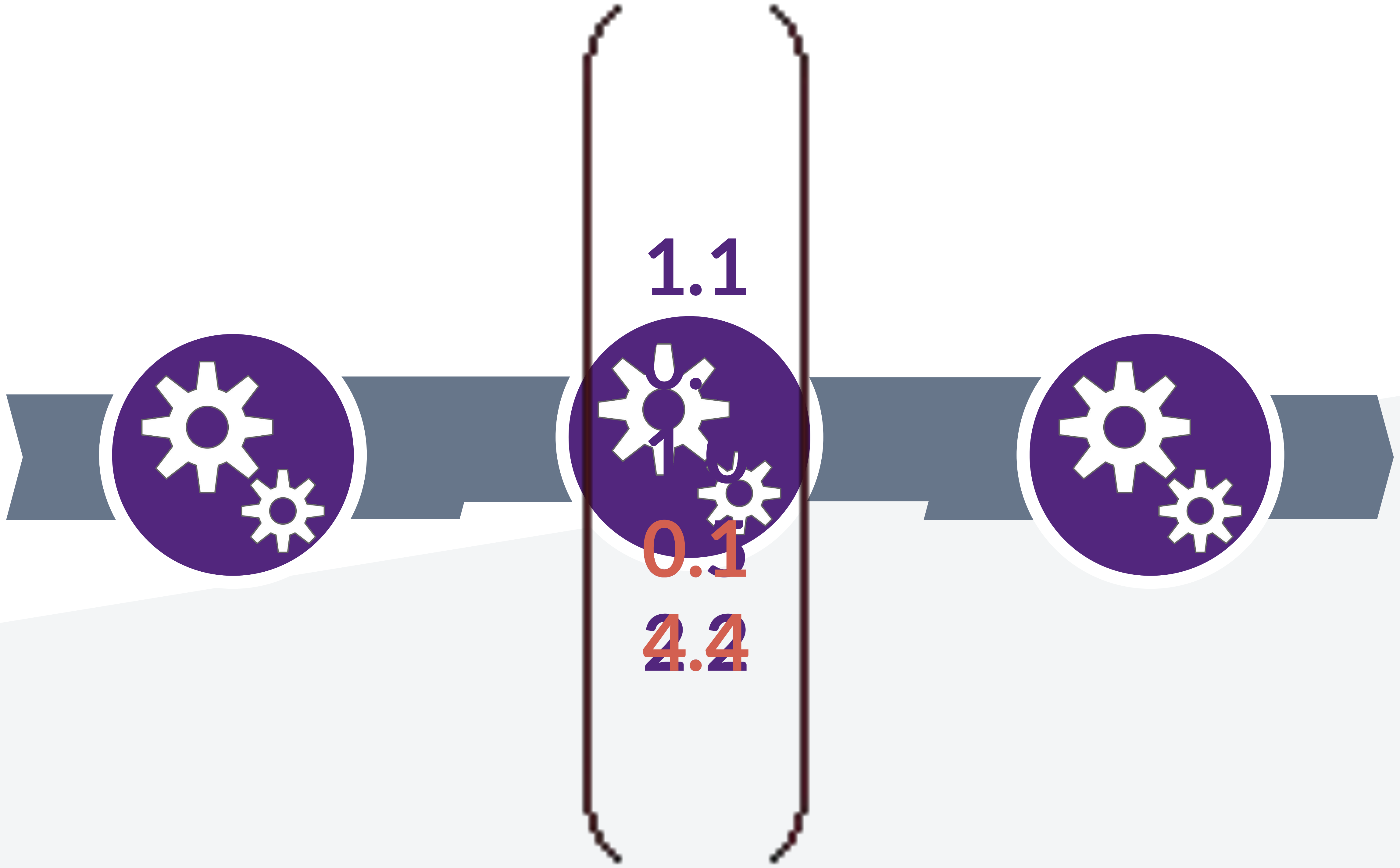
MultiPie

- 750,000 images
- 337 people
- 19 illuminations
- 15 camera views
- 6 expressions



 http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Content.html

InstFateLglyeFace



Face embedding

How to create this Face Embedding?

Auto-encoder

Encoder

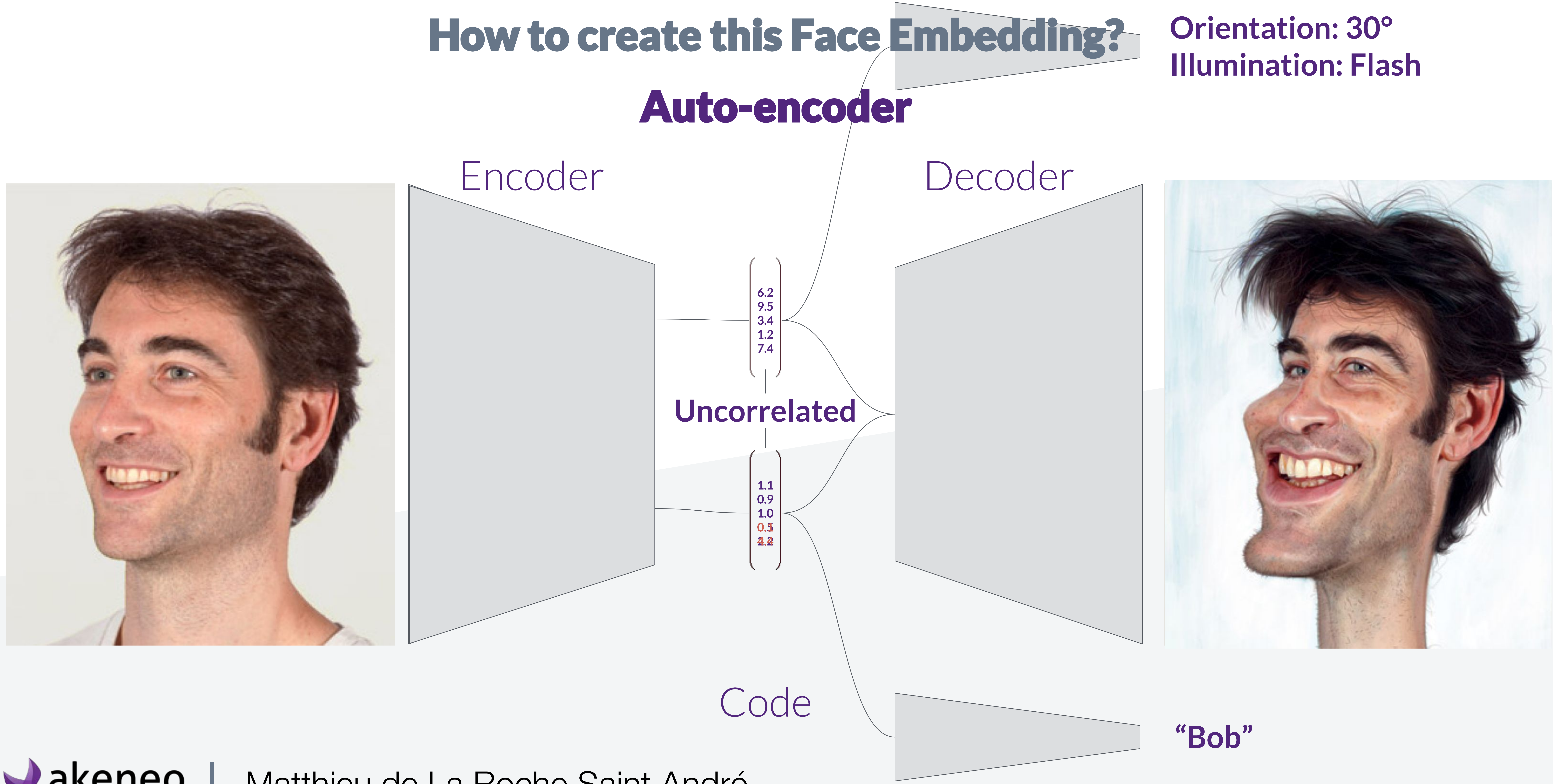
1.1
0.9
1.0
0.5
2.2

Code

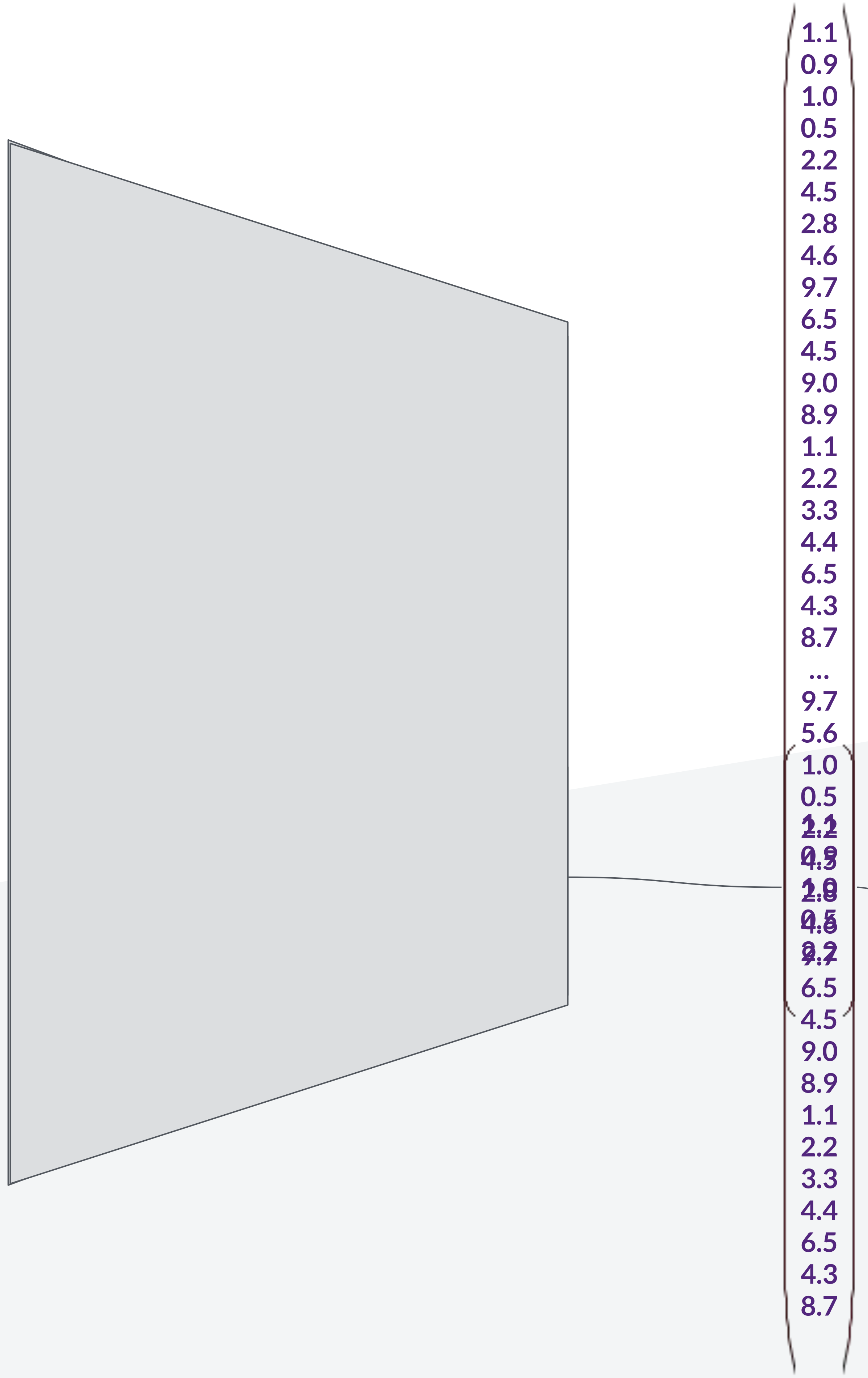
Decoder

How to create this Face Embedding?

Auto-encoder

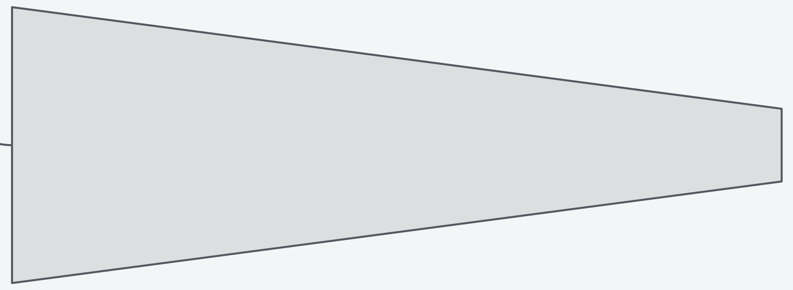


Classification Task



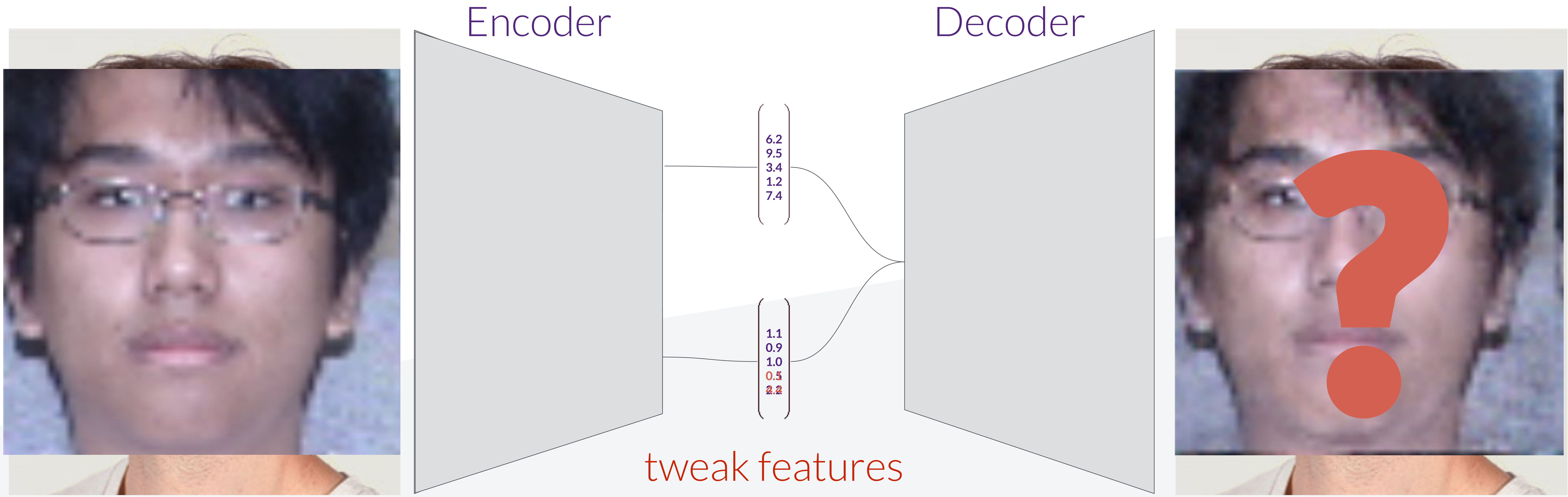
How many features are necessary?
Fewer and higher-level features is better

Face Embedding Size	Accuracy
1024	100.00%
200	99.76%
50	99.76%
10	78.78%



“Bob”

Reconstruction Task

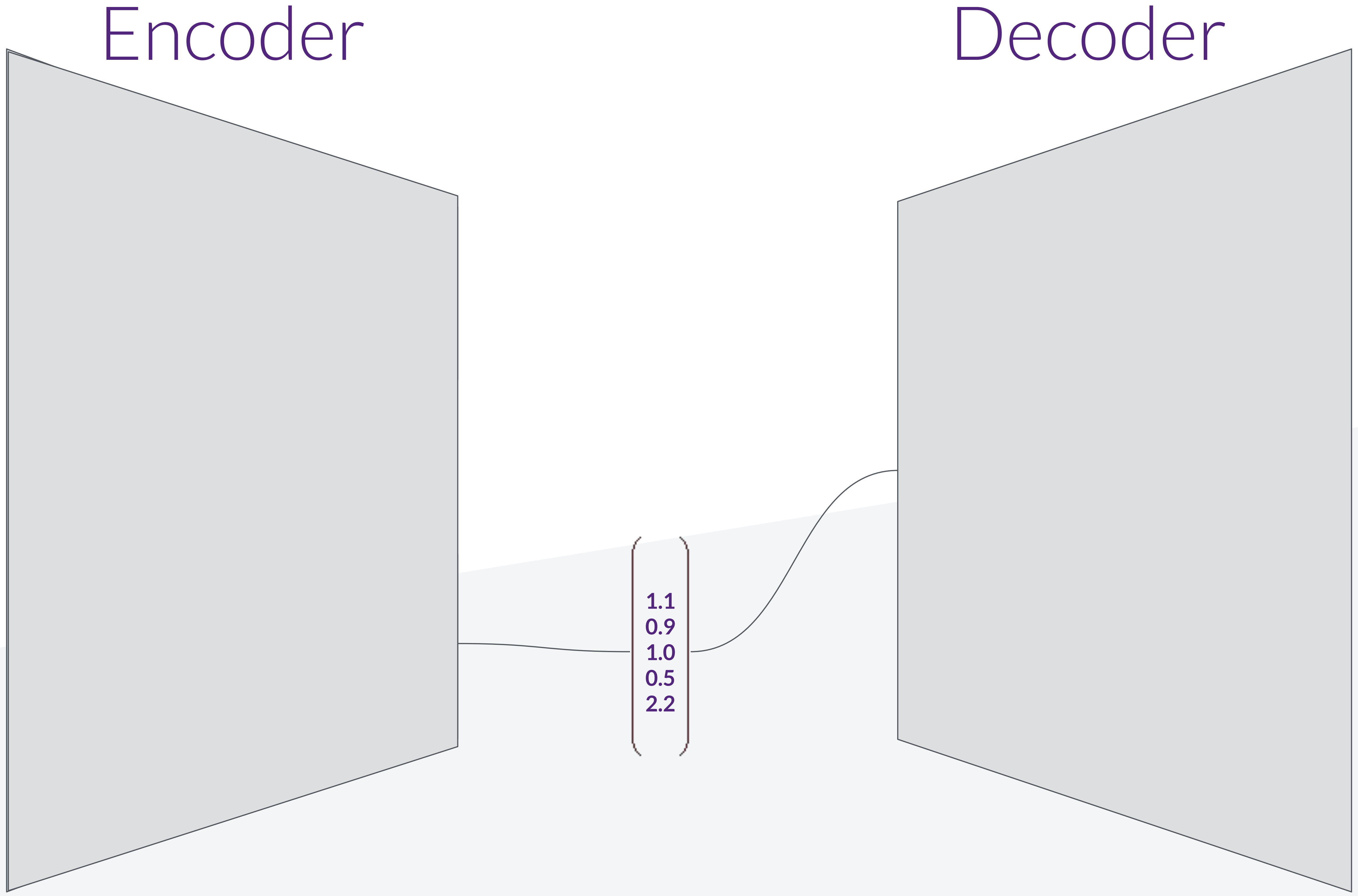


Difference between features and average features has been multiplied by **2**

Caricature Generation

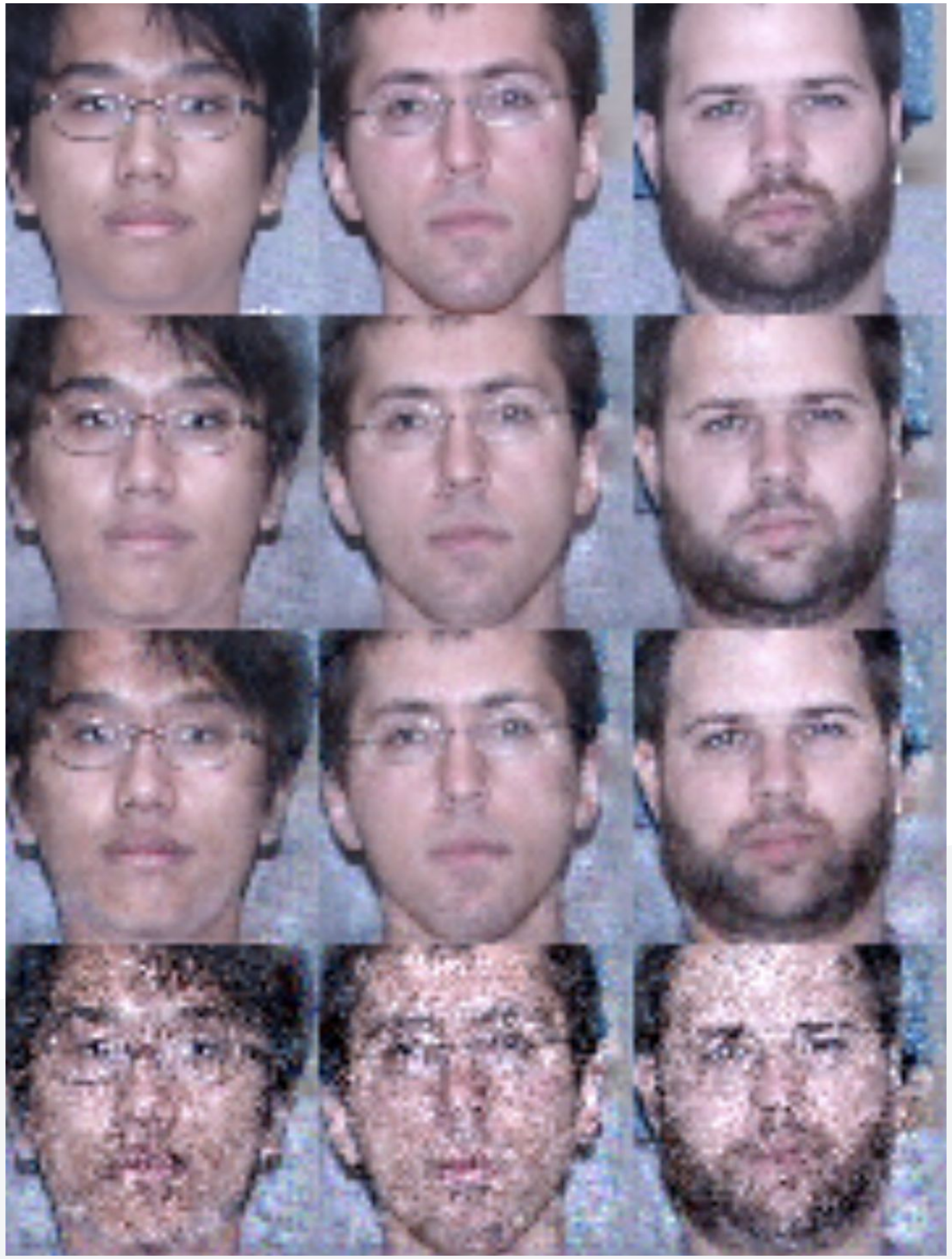
Difference between features and average features has been multiplied by **5**

Reconstruction With Less Features



Reconstruction Task

New Faces (ie Test Set)

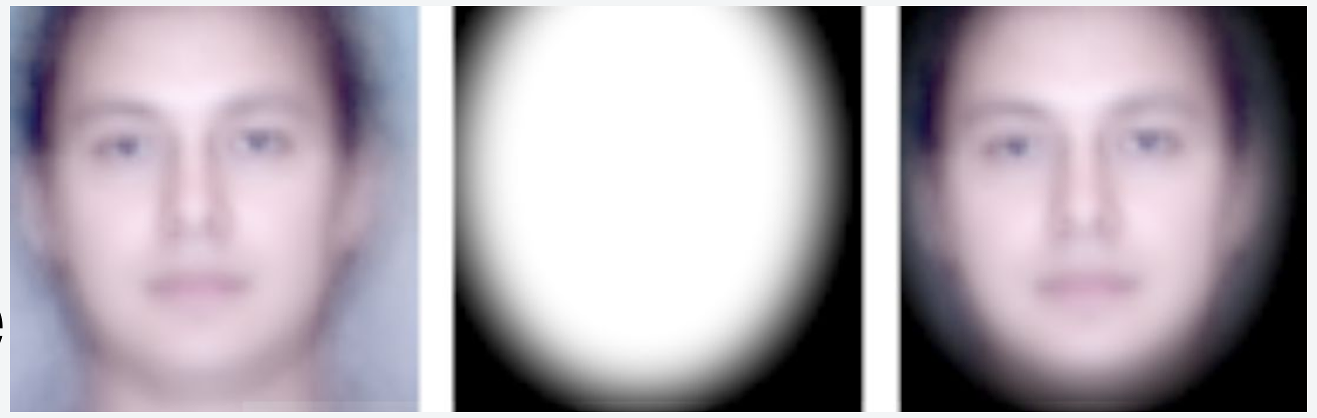


Original Image

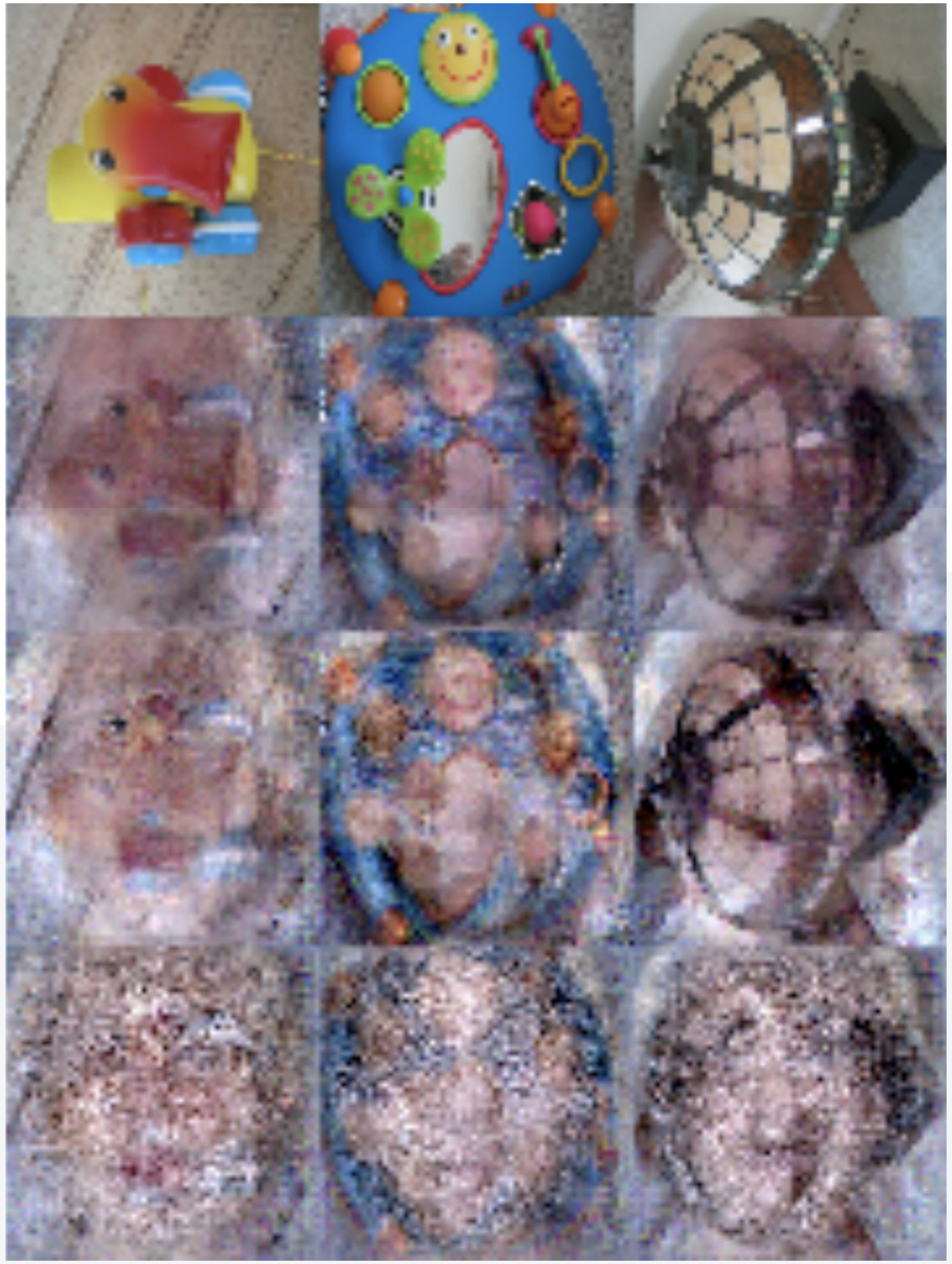
Face Embedding : 50 floats

Face Embedding : 1 floats

Face Embedding : **0** float



New and Unrelated Images

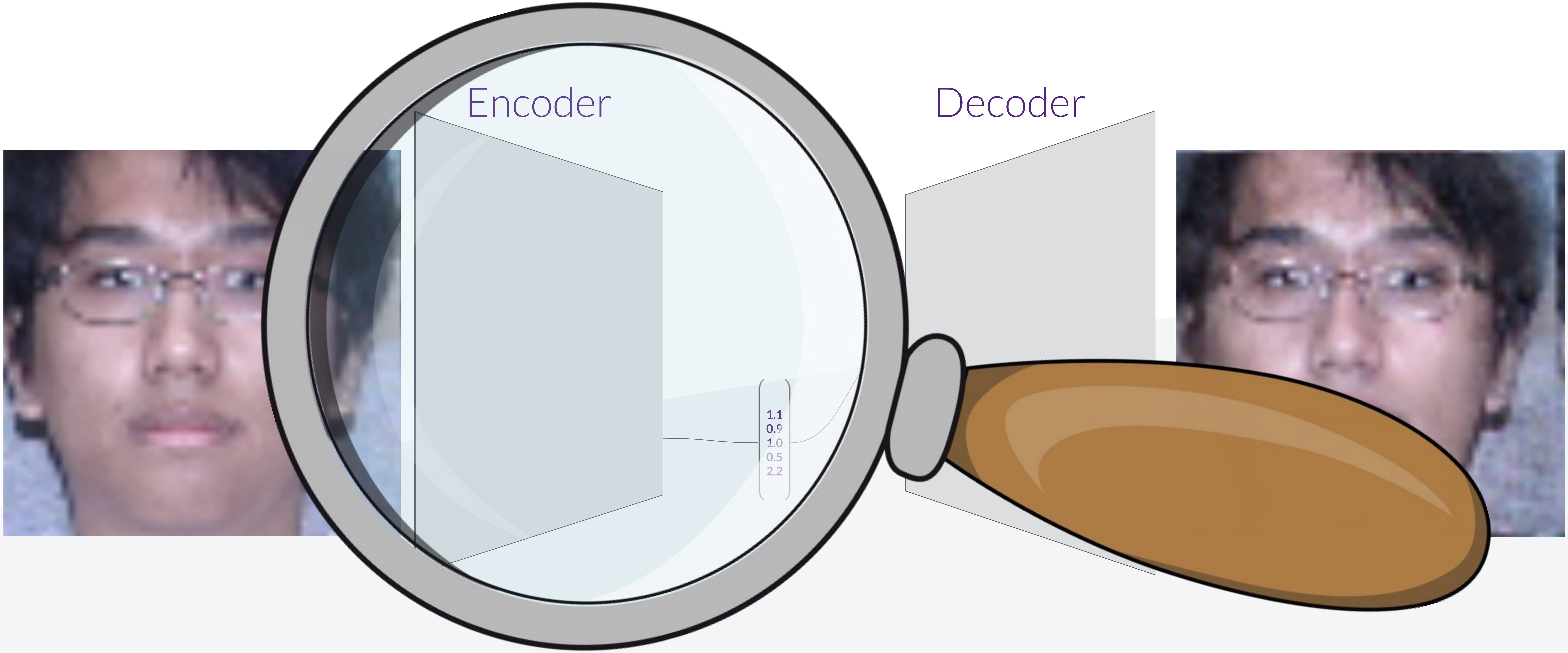


Overfitting on faces?

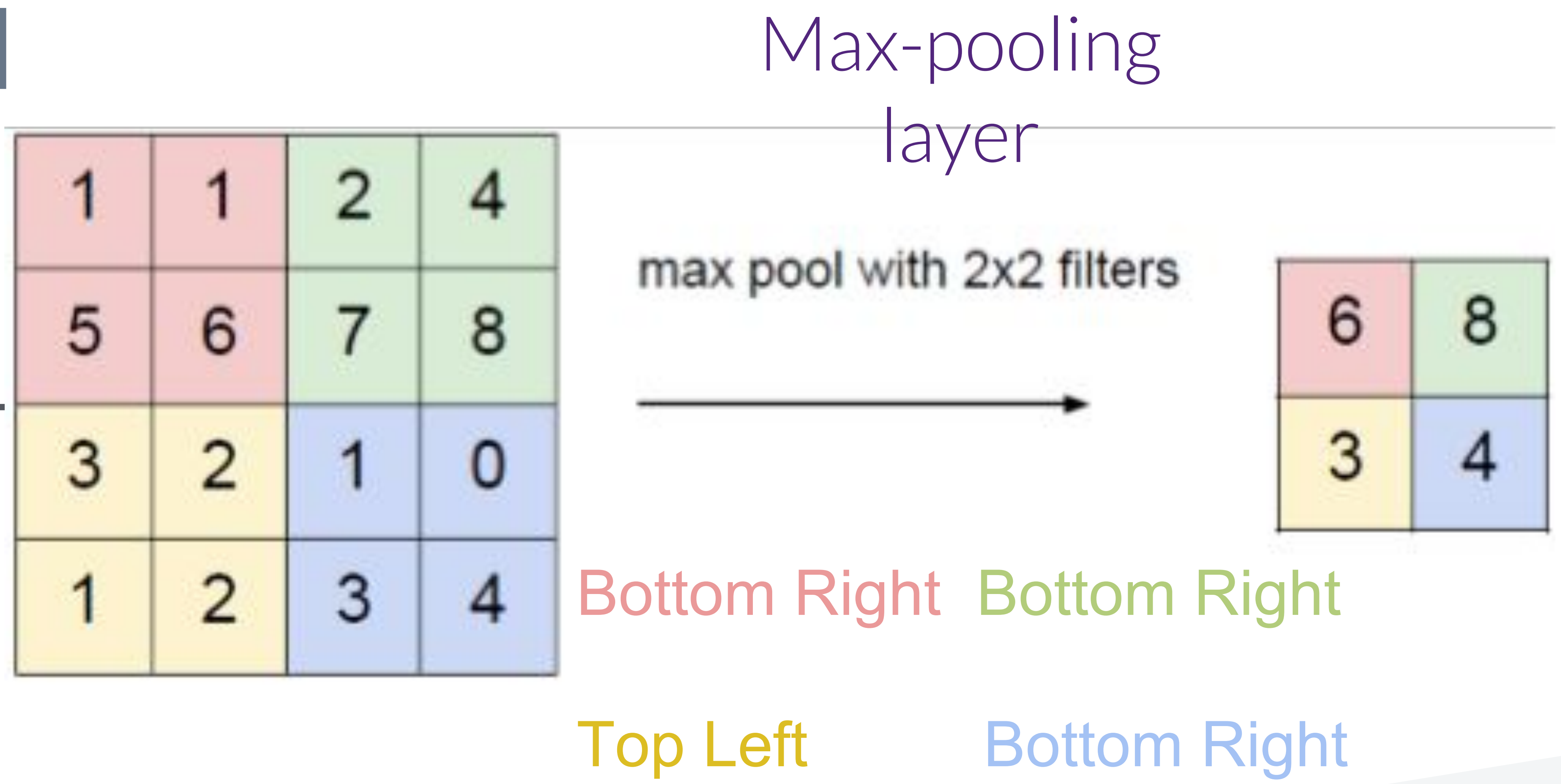
> Tunnel Effect <

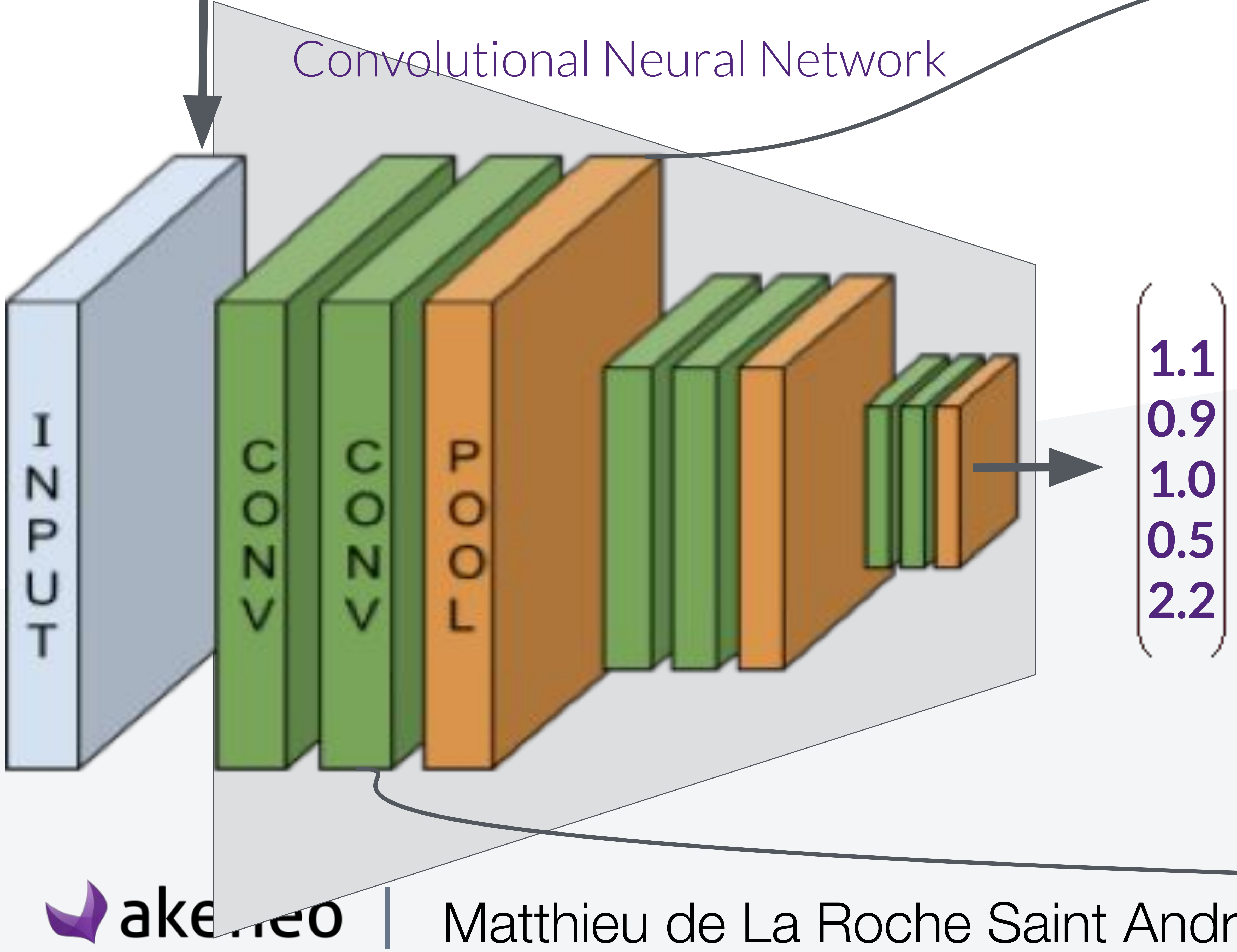
Tunnel Effect
Why?
How did this happen?

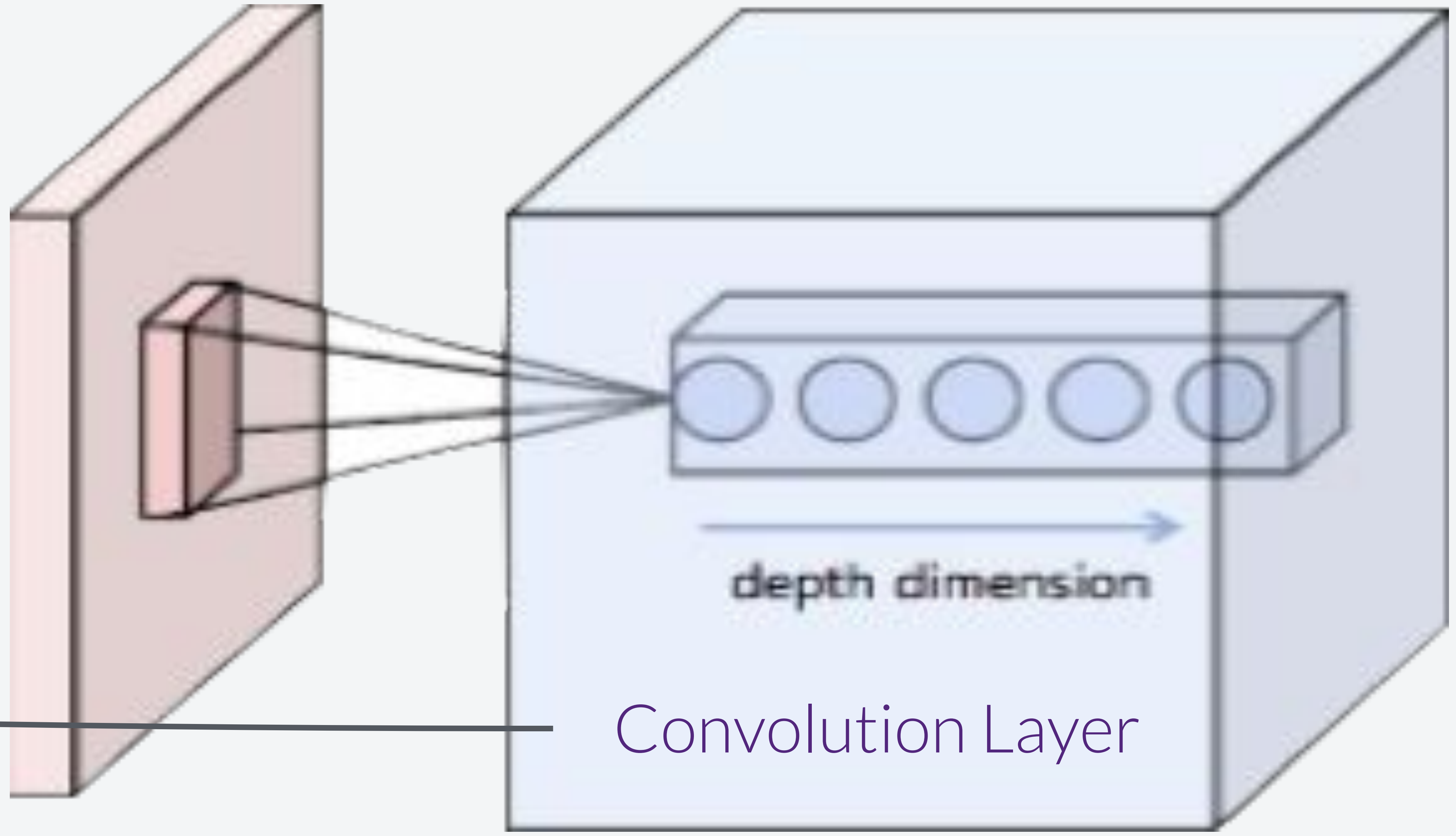
Diving into the network



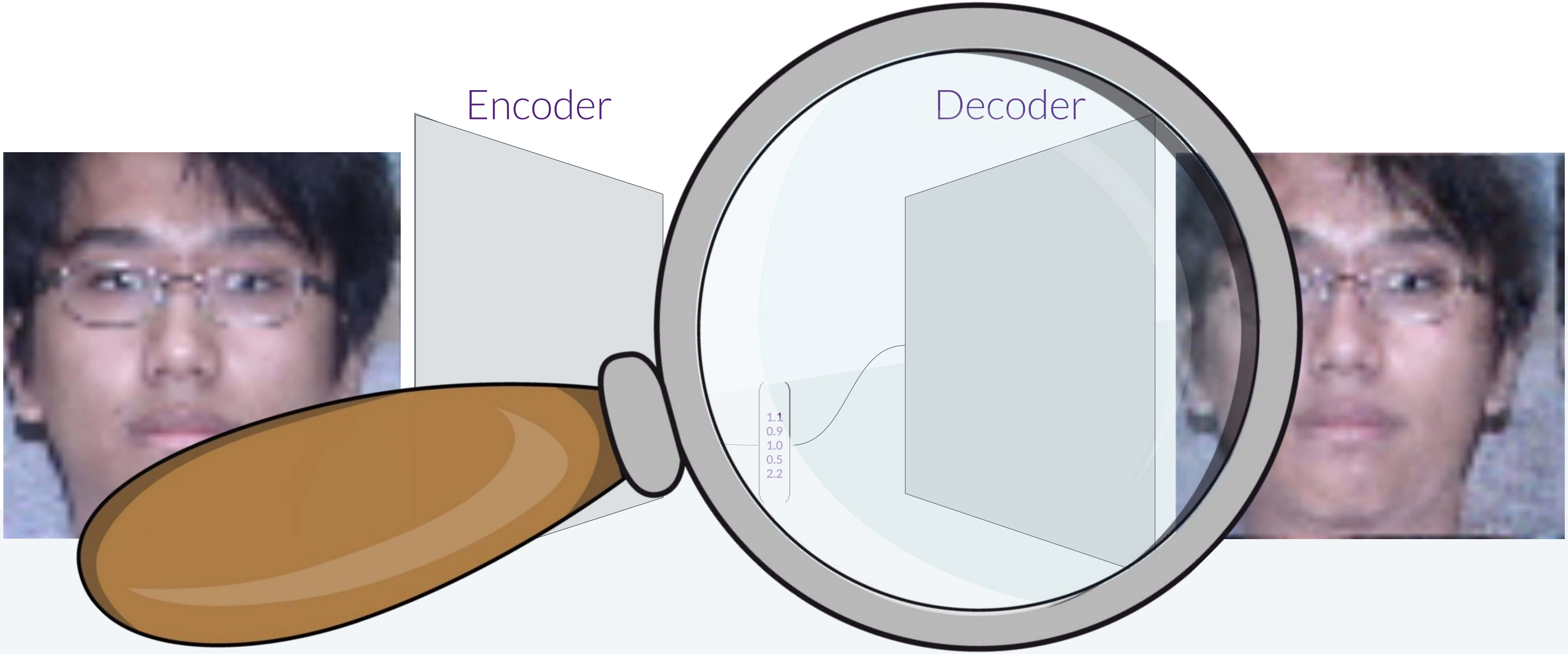
The Encoder: CNN





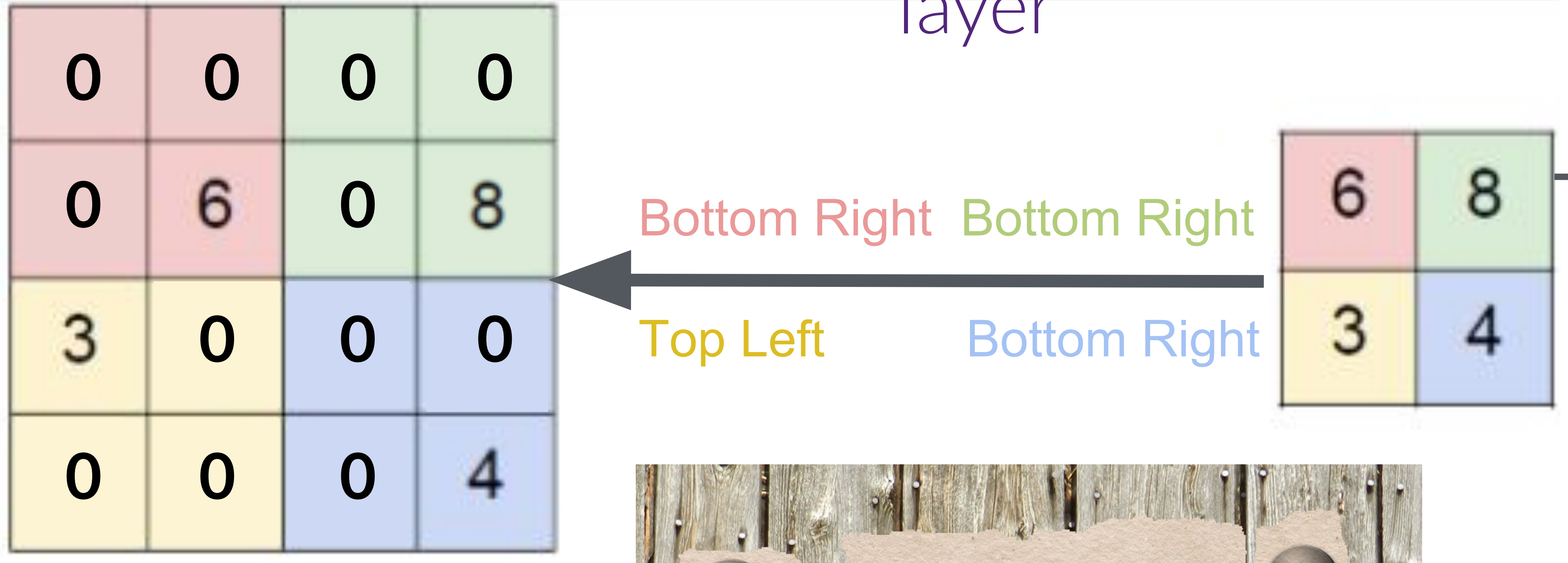


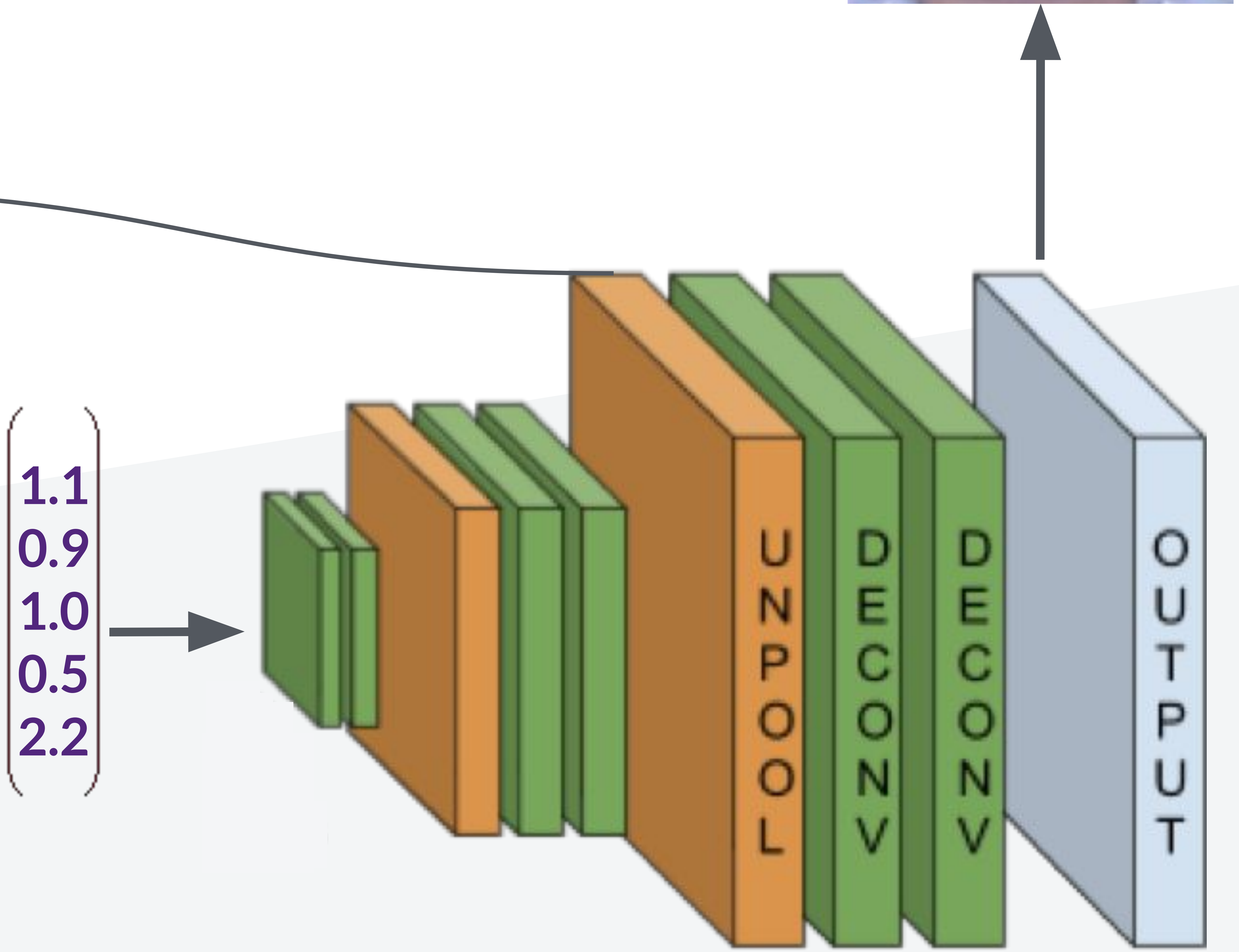
Diving into the network



The decoder

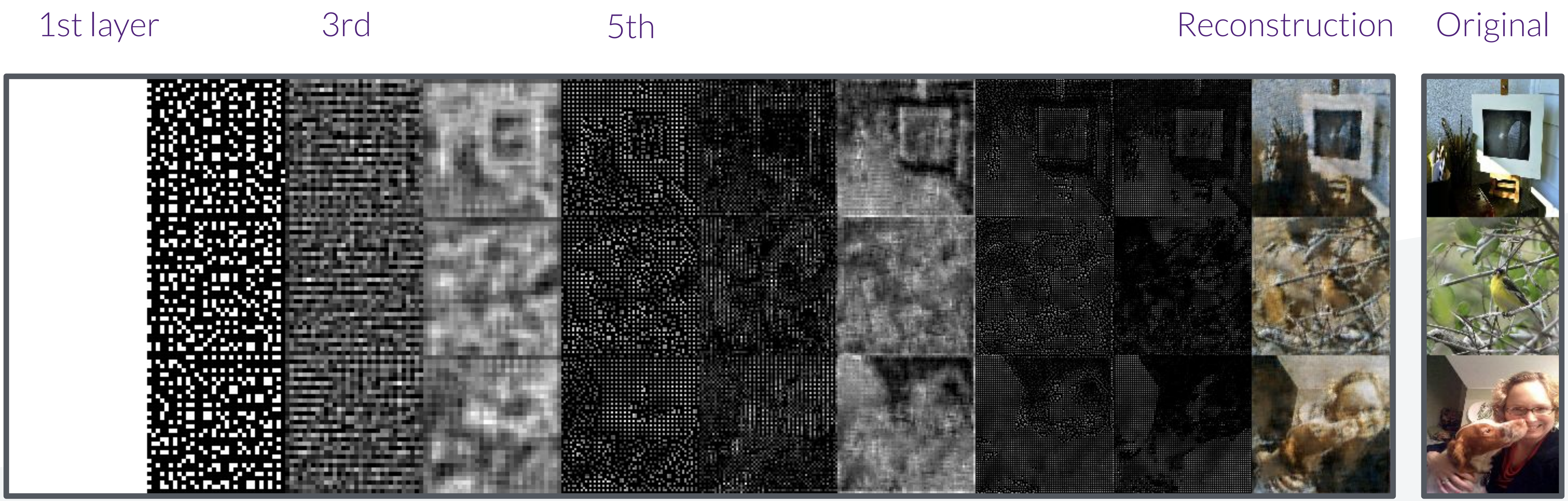
Max-unpooling layer



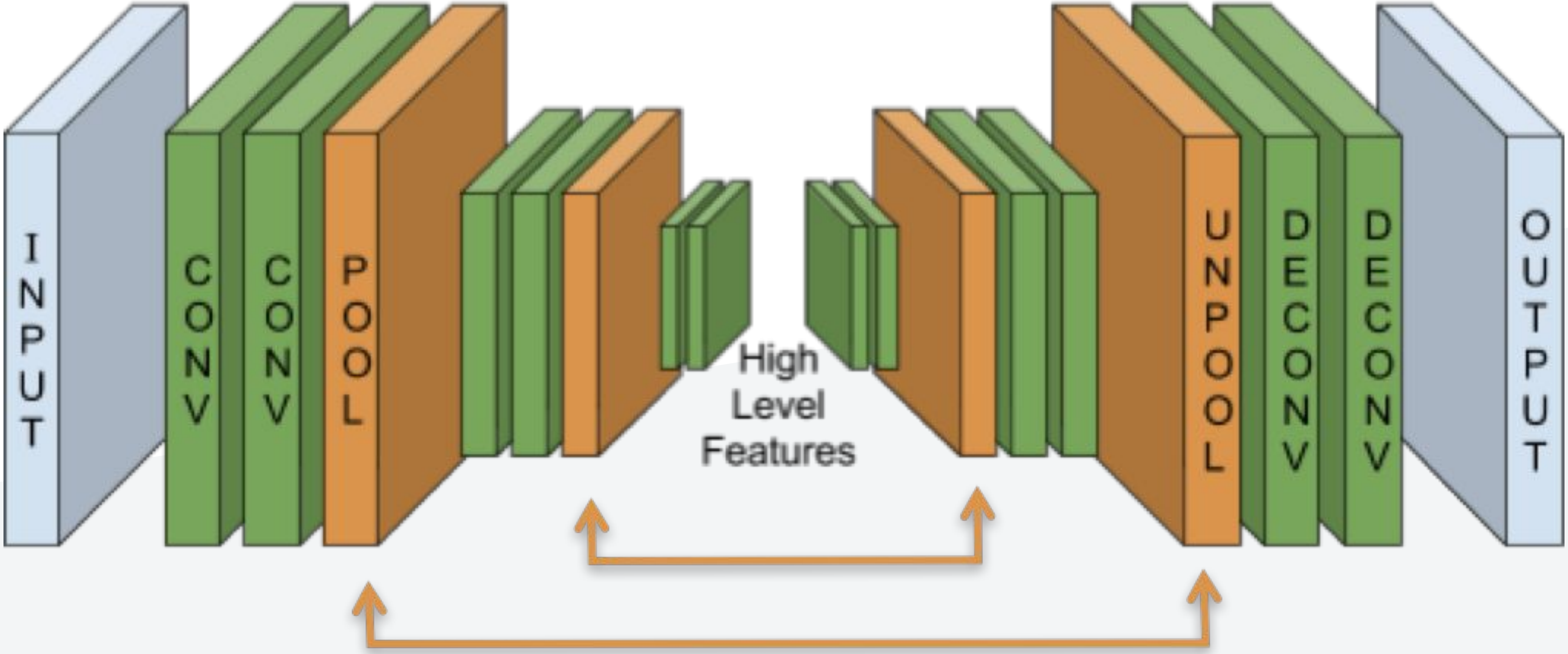


The decoder

$$\begin{pmatrix} 1.1 \\ 0.9 \\ 1.0 \\ 0.5 \\ 2.2 \end{pmatrix}$$

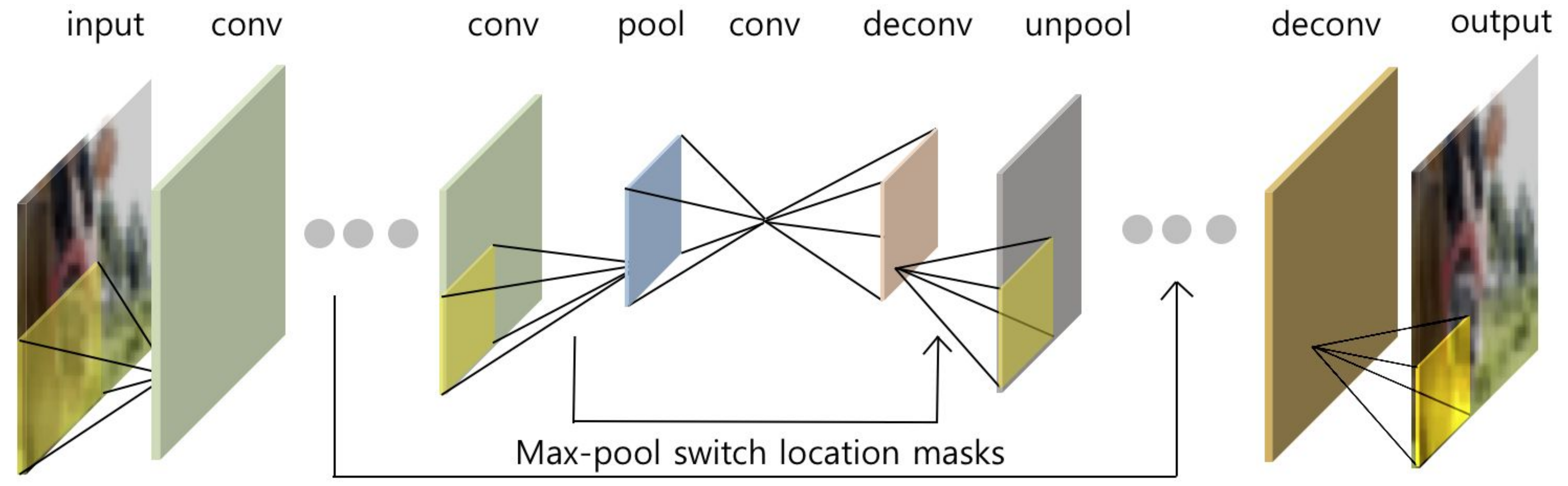


Leak: Max-pool switch location mask



Max-pool switch location mask
2-14 bits / pixel

Leak:
Max-pool switch location mask



Random Encoder

- Original images

- Encoder with random weights
- (untrained) [16,16,16]

- Encoder pre-trained
- for classification [16,16,16]

- Encoder with random weights
- (untrained) [48,96,192]

Random Encoder

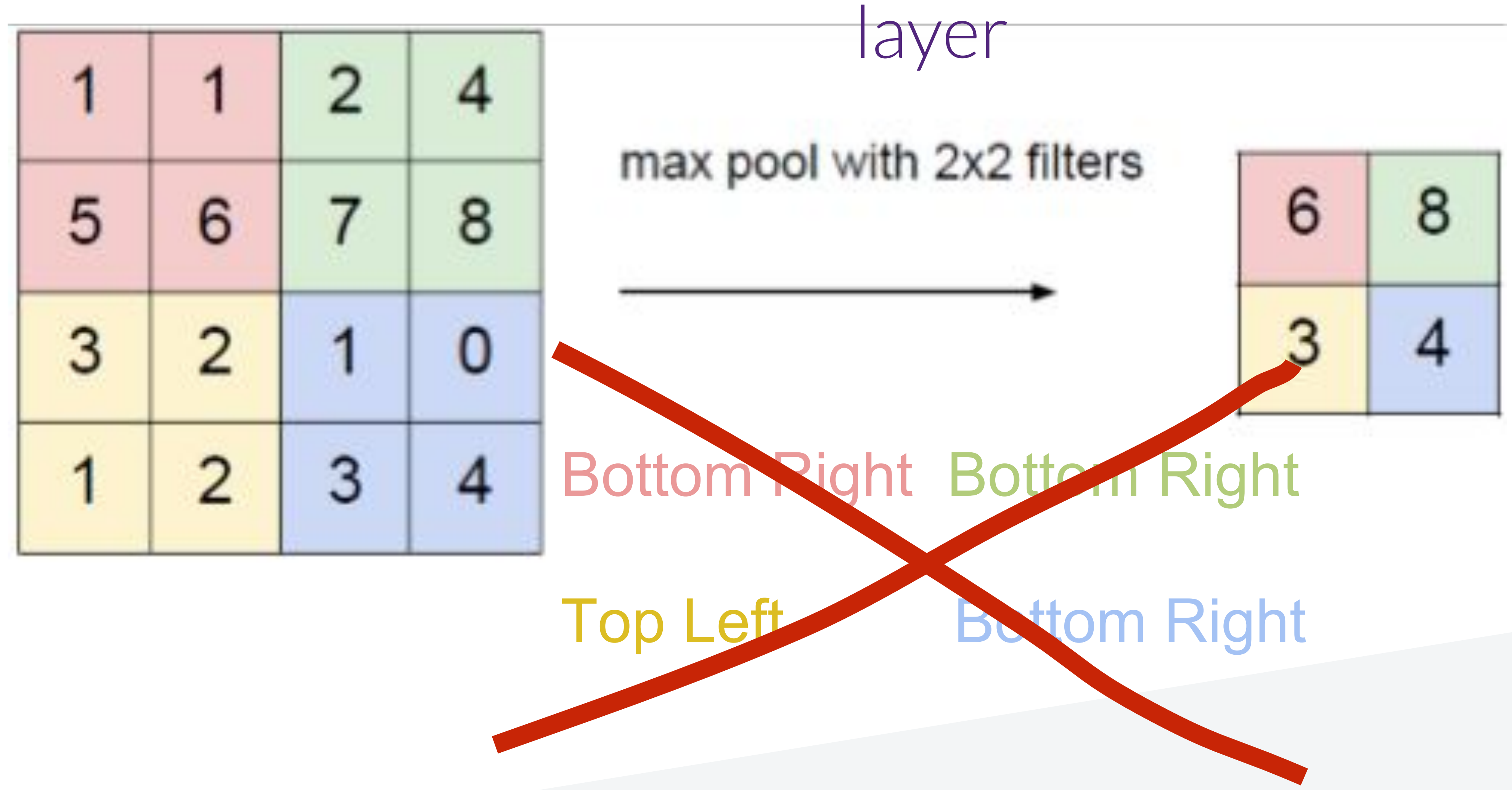
Root Mean Square Error

TABLE II
RMSE AND BITS USED FOR ALL ARCHITECTURES

Architecture	Reconstruction RMSE test [pixels in range [0;1]]		Max-switch information [bits per pixel]
Pretrained	No	Yes	
[8,8,8]	0.1464	0.1545	1.75
[6,12,24]	0.1387	0.1464	1.75
[16,16,16]	0.1210	0.1156	3.5
[12,24,48]	0.1138	0.1206	3.5
[32,32,32]	0.0922	0.0950	7
[24,48,96]	0.0876	0.0915	7
[64,64,64]	0.0744		14
[48,96,192]	0.0718		14

Bonus Track #1

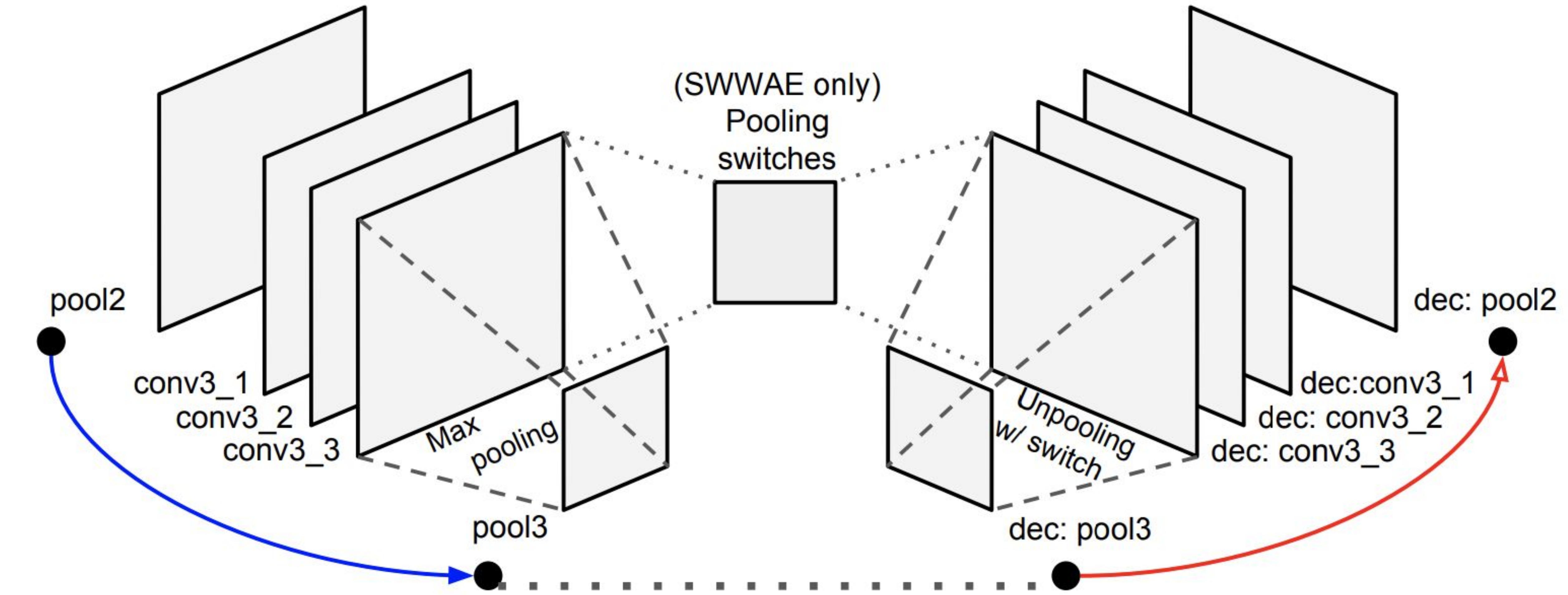
Max-pooling layer



Average-pooling doesn’t work at all

Bonus Track #2

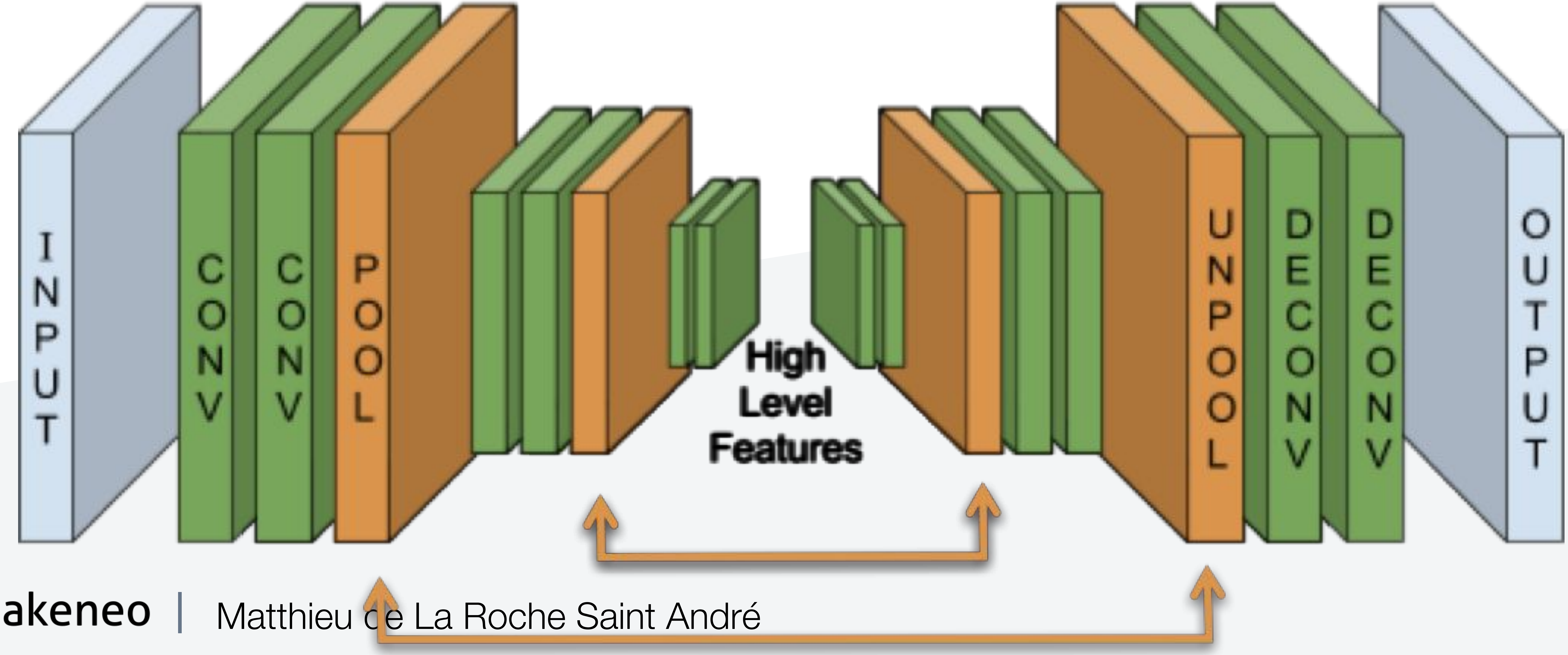
Y. Zhang, K. Lee, and H. Lee, "Augmenting supervised neural networks with unsupervised objectives for large-scale image classification," ICML, pp. 612–621, 2016.



& Brox (2016), we use the auxiliary decoding pathway of the stacked autoencoder to reconstruct images from intermediate activations of the pretrained classification network. Using SWWAE, we demonstrate better image reconstruction qualities compared to the autoencoder using the unpooling operators with *fixed* switches, which upsamples an activation to a fixed location within the kernel. This result suggests that the intermediate (even high-level) feature representations preserve nearly all the information of the input images except for the locational details "neutralized" by max-pooling layers.

Bonus Track #3

Oreo Training



Take away

- Images reconstructed from *max-pool switch location mask*
- 🎵 Mind the leak between the encoder and the decoder 🎼 🚄
- Potential pitfall for future architectures
- Don’t even need to train the encoder
- Neural networks are amazingly adaptable
- Potential usages:
  - image compression
  - *max-pool switch location mask* as a new image space representation
  - single-forward pass artistic style transfer
  - improve image segmentation

Questions?

Tunnel Effect in CNNs: Image Reconstruction From Max-Switch Location

Matthieu de La Roche Saint Andre, Laura Rieger, Morten Hannemose, and Junmo Kim, "Tunnel Effect in CNNs: Image Reconstruction From Max-Switch Locations," IEEE Signal Processing Letters, vol. 24, no. 3, pp. 254–258, Mar. 2017.
http://ieeexplore.ieee.org/document/7781571/
http://orbit.dtu.dk/ws/files/128168371/double.pdf