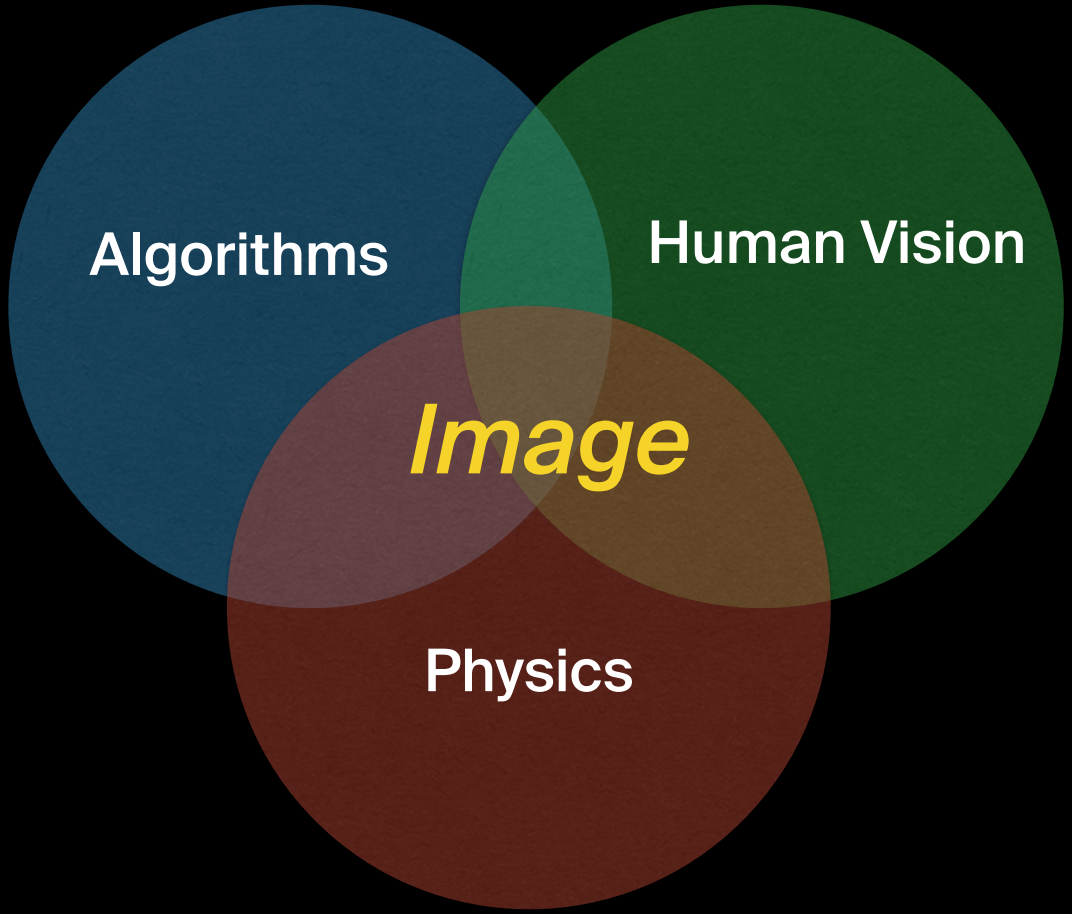
Uncovering local semantics in CNNs and GANs

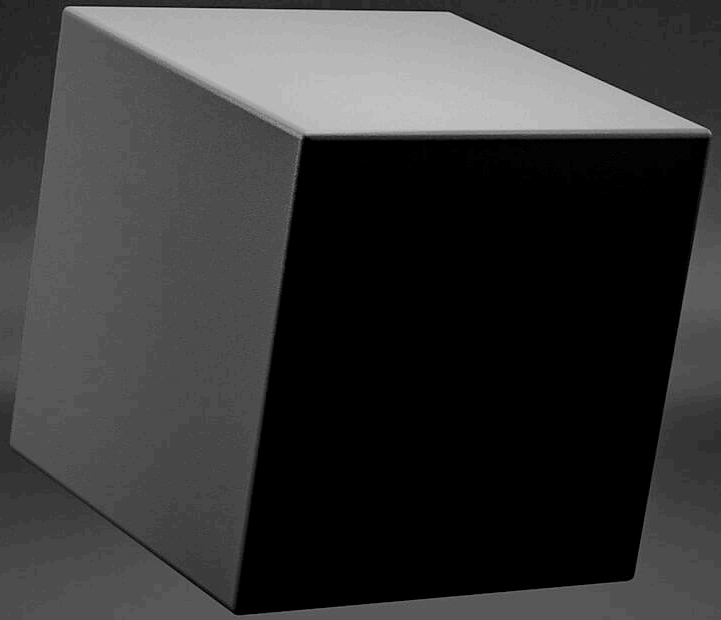
Sabine Süsstrunk

IVRL/IC/EPFL

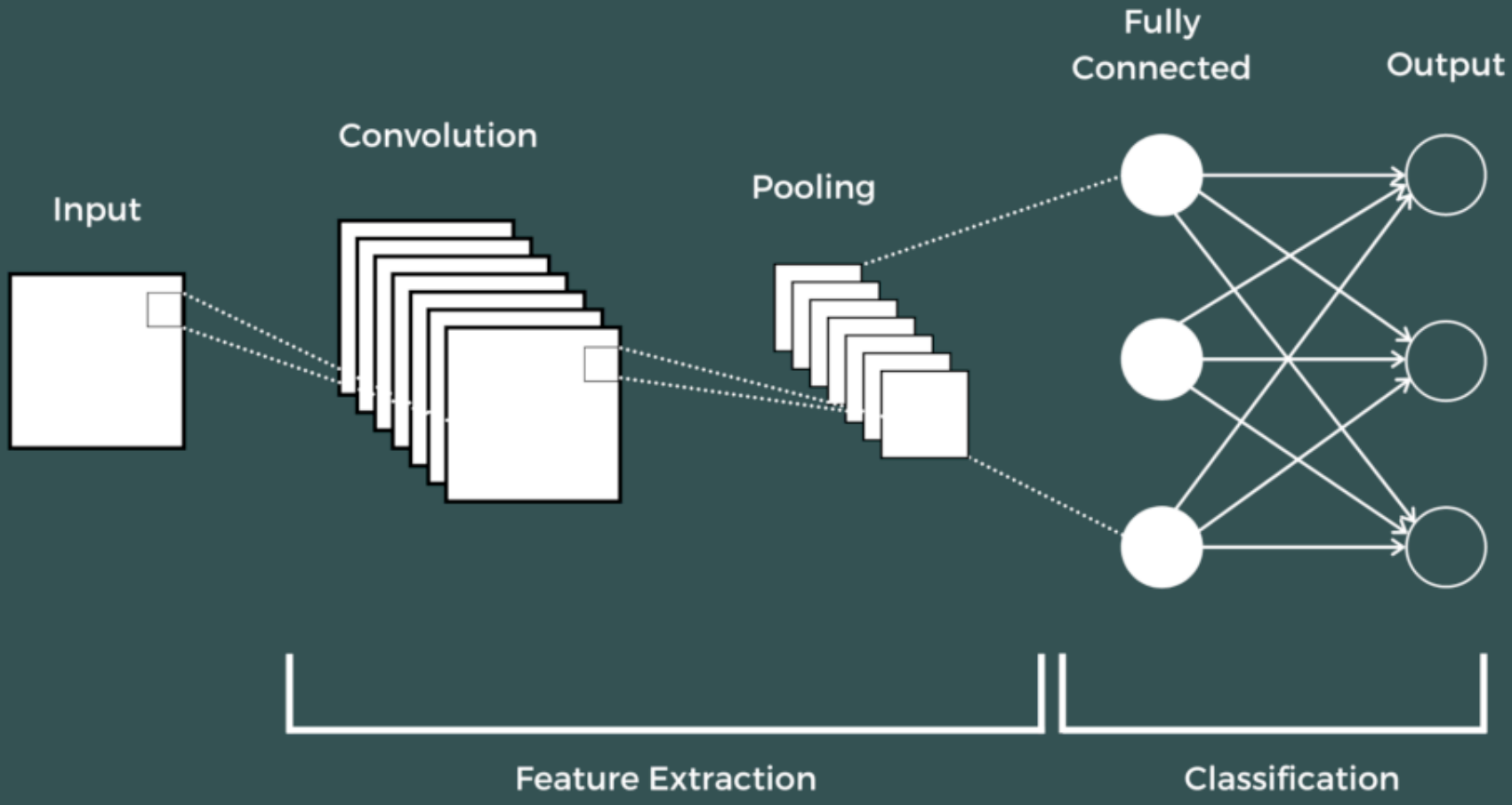
EPFL IVRL



Copyright: Juan Carrano

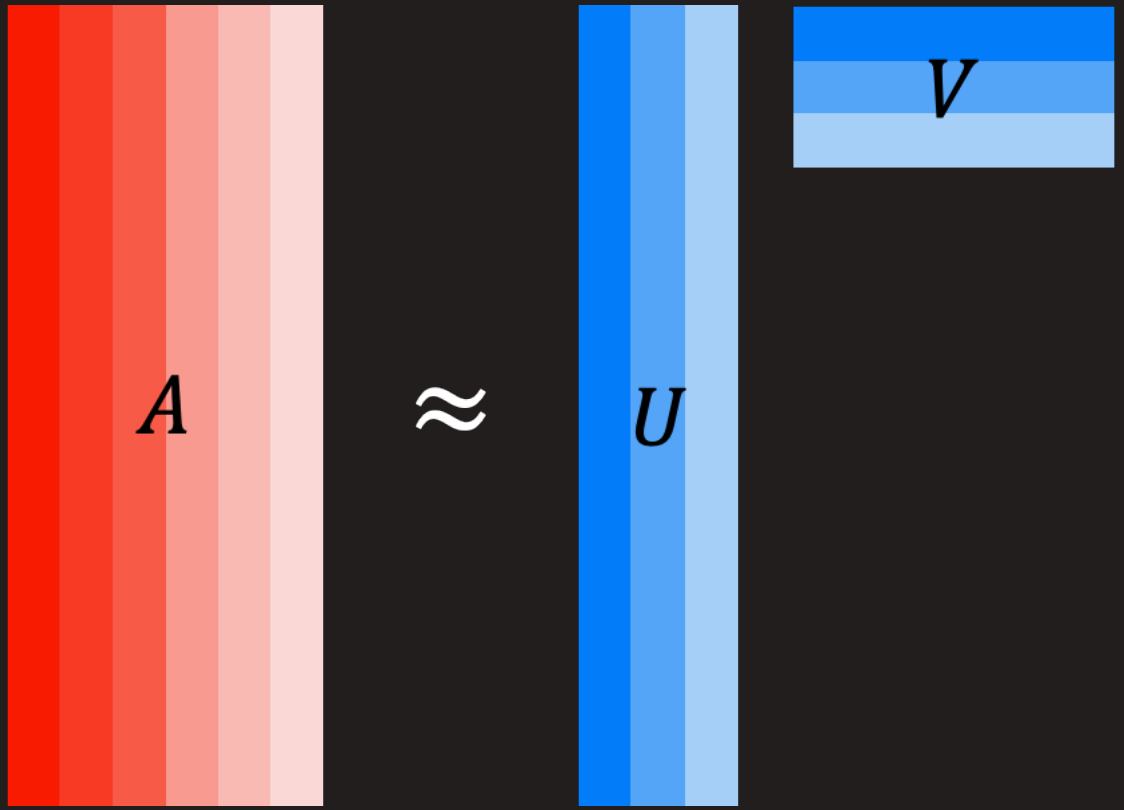


https://singularityhub.com/2019/04/17/in-defense-of-black-box-ai/

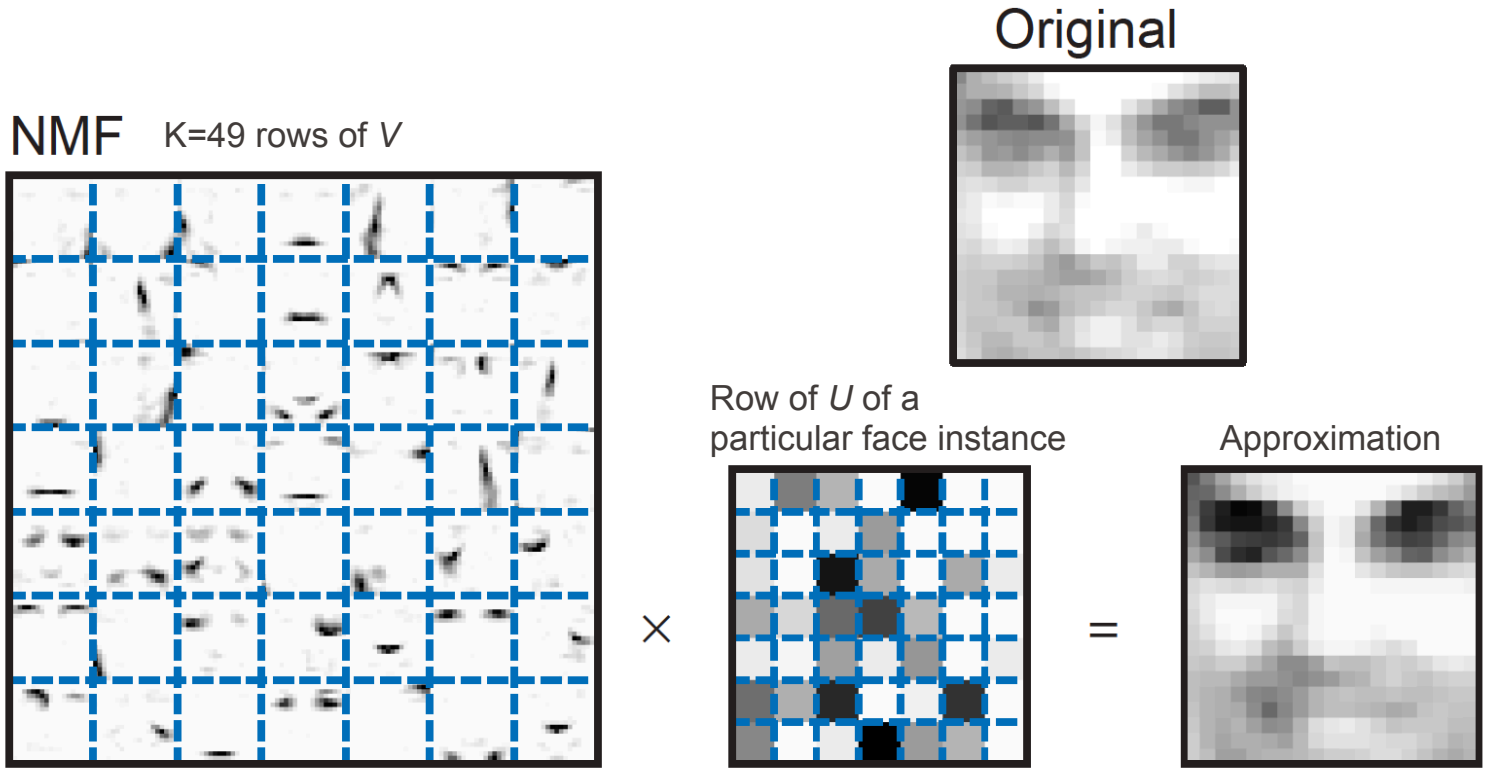


https://www.theclickreader.com/building-a-convolutional-neural-network/

Matrix Factorization



Previous Work



D. D. Lee and H. S. Seung, Learning the Parts of Objects by Non-Negative Matrix Factorization" *Nature,* 1999.

Non-Negativ Matrix Factorization with ReLU based CNN's

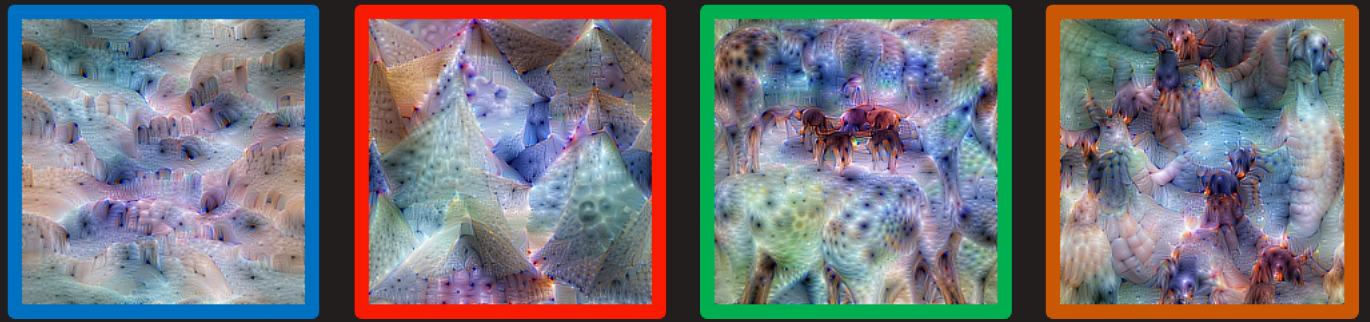
$$\max(0,\cdot) \longrightarrow \underbrace{A}_{R_+^{P\times M}} \approx \underbrace{U}_{R_+^{P\times K}} \underbrace{V}_{R_+^{K\times M}}$$

ReLU

Edo Collins, Radhakrishna Achanta, Sabine Süsstrunk, Deep Feature Factorization for Concept Discovery, ECCV 2018

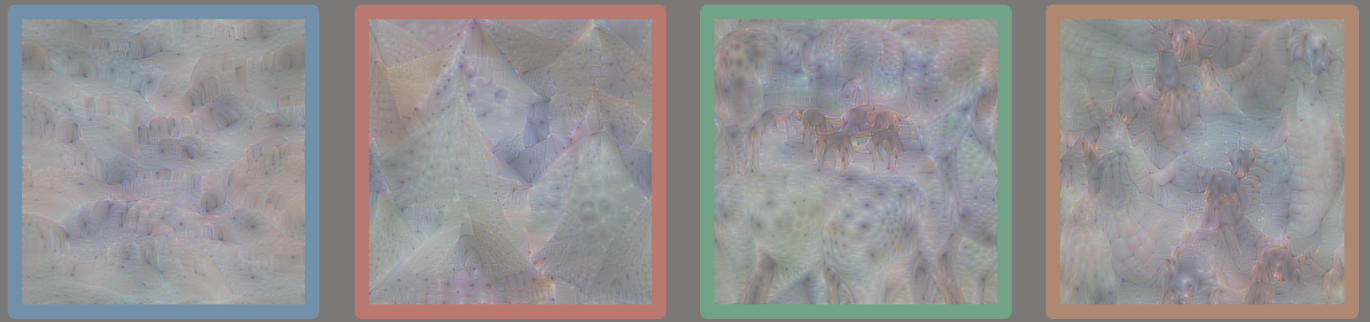
Semantic localization with matrix *U:* where

Content-based image retrieval with matrix *V:* what



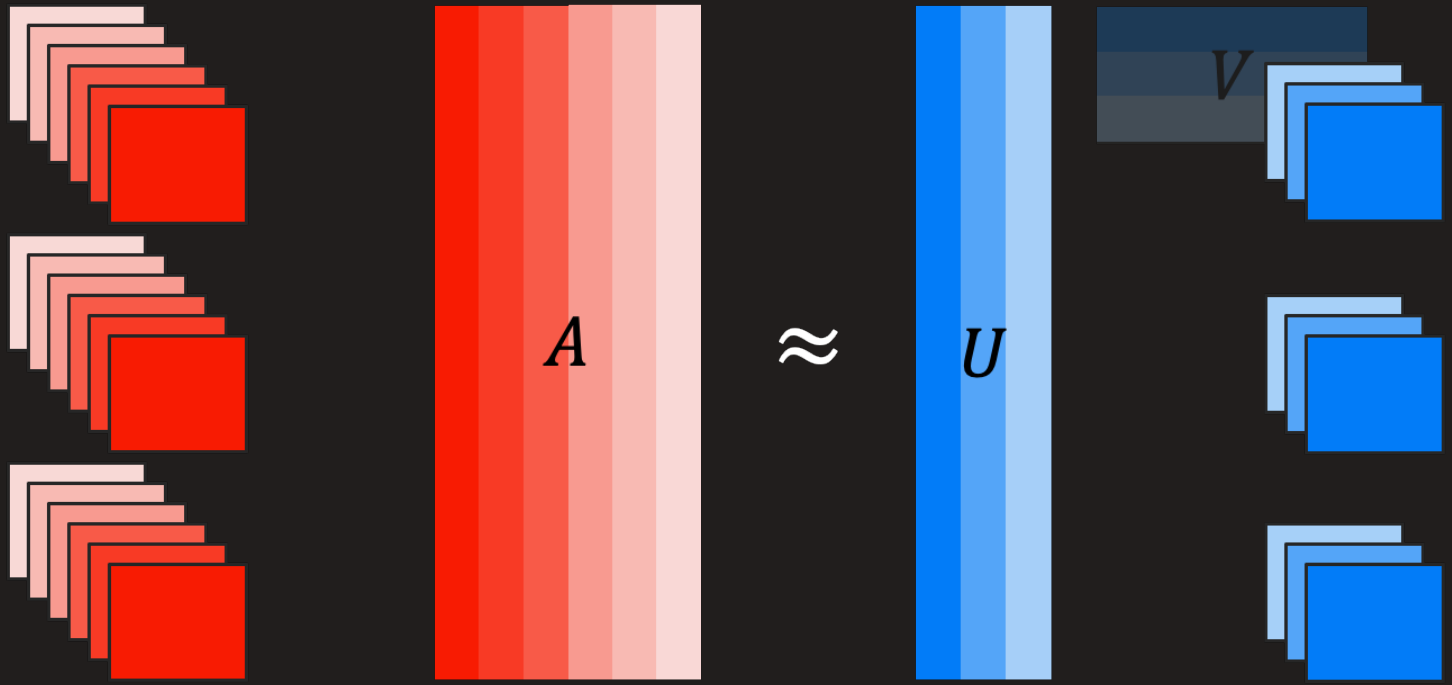
Semantic localization with matrix *U:* where

Content-based image retrieval with matrix *V:* what



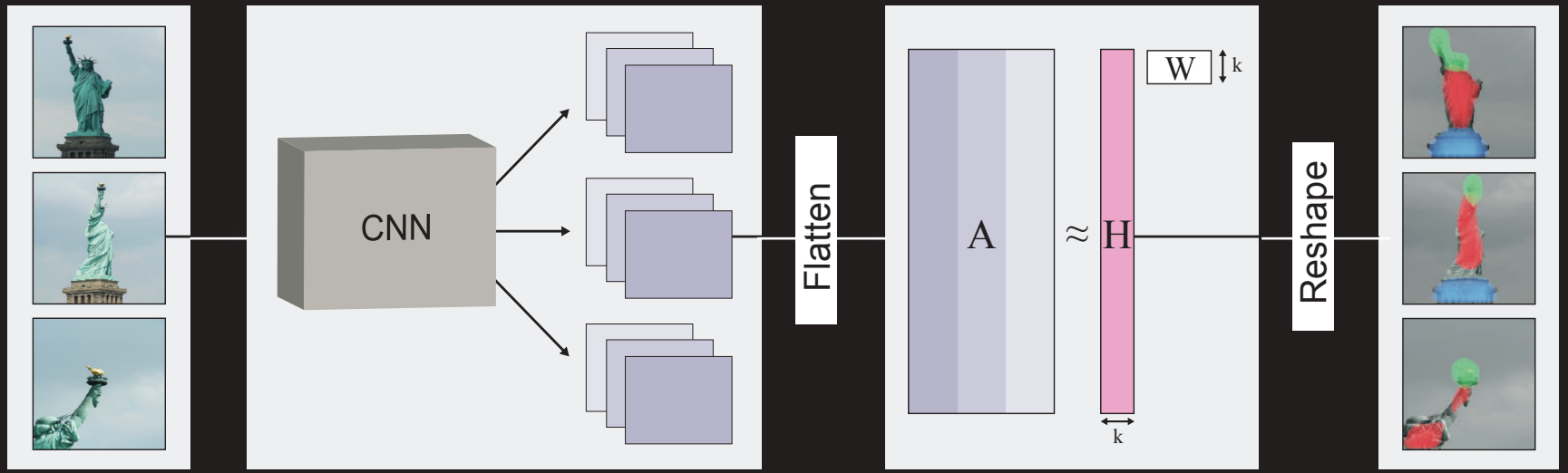
Reshaping U into heatmaps

$$N \times C \times H \times W \rightarrow (N \cdot H \cdot W) \times C$$



$$(N \cdot H \cdot W) \times K \rightarrow N \times K \times H \times W$$

Deep Feature Factorization



Edo Collins, Radhakrishna Achanta, Sabine Süsstrunk, Deep Feature Factorization for Concept Discovery, ECCV 2018

NMF heatmaps – VGG-19

$K = 1$

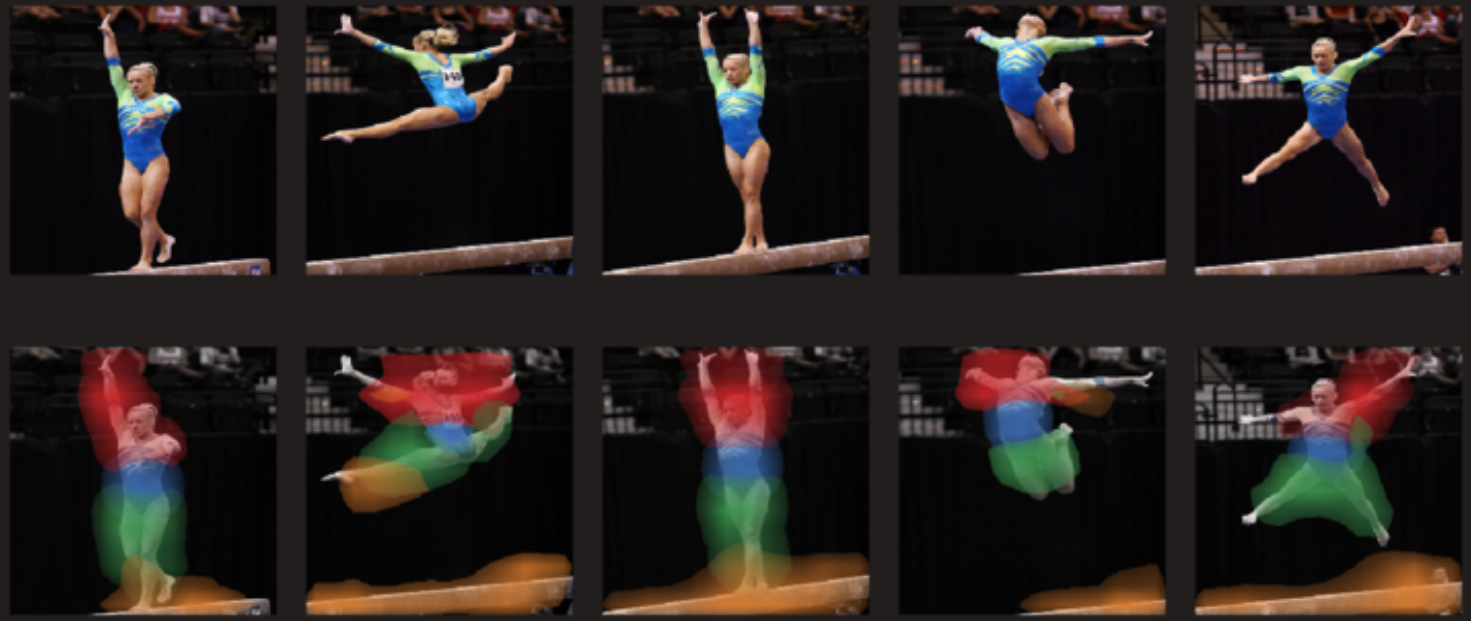
NMF heatmaps – VGG-19

$K = 2$

NMF heatmaps – VGG-19

$K = 3$

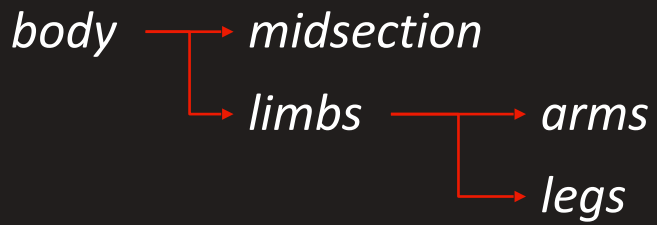
NMF heatmaps – VGG-19



$K = 4$

Semantic decomposition

- For VGG-19, K=1 often highlights the salient object.
- Setting K>1 results in a decomposition into semantic parts.
- Incrementing K reveals a cluster hierarchy in feature space, e.g.



NMF heatmaps – VGG-19

$K = 1$

NMF heatmaps – VGG-19

$K = 2$

NMF heatmaps – VGG-19

$K = 3$

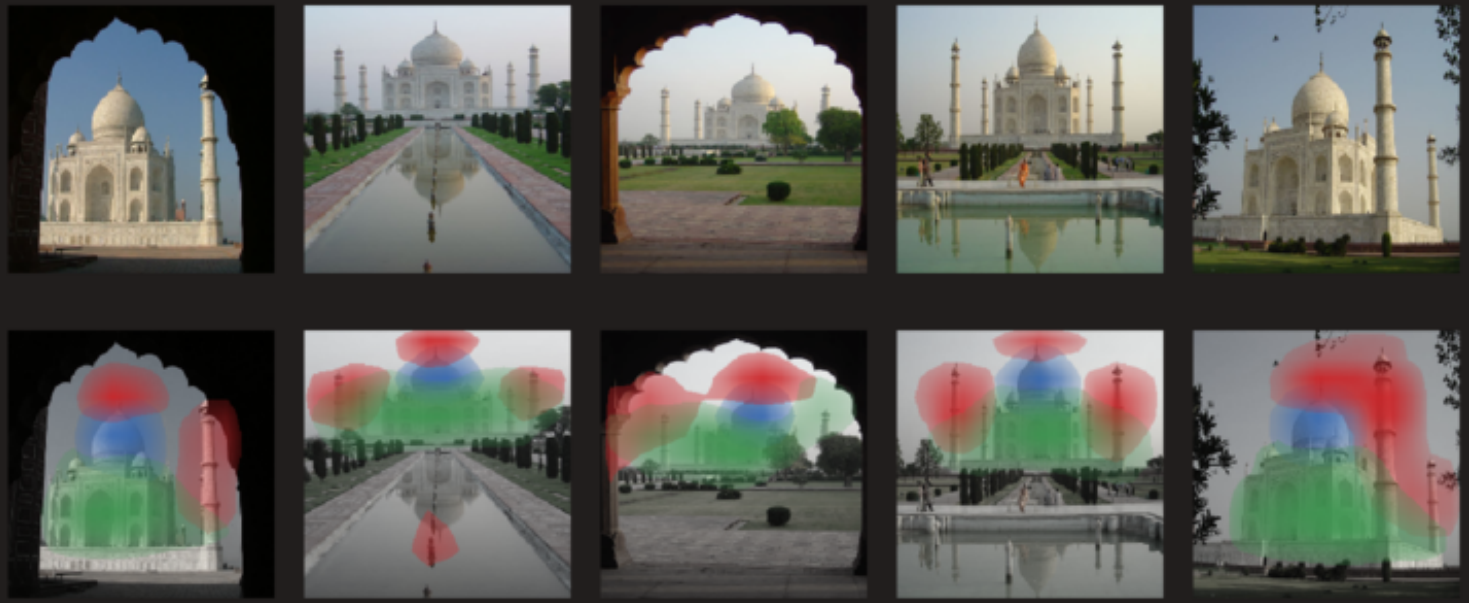
NMF heatmaps – VGG-19

$K = 4$

NMF heatmaps – VGG-19

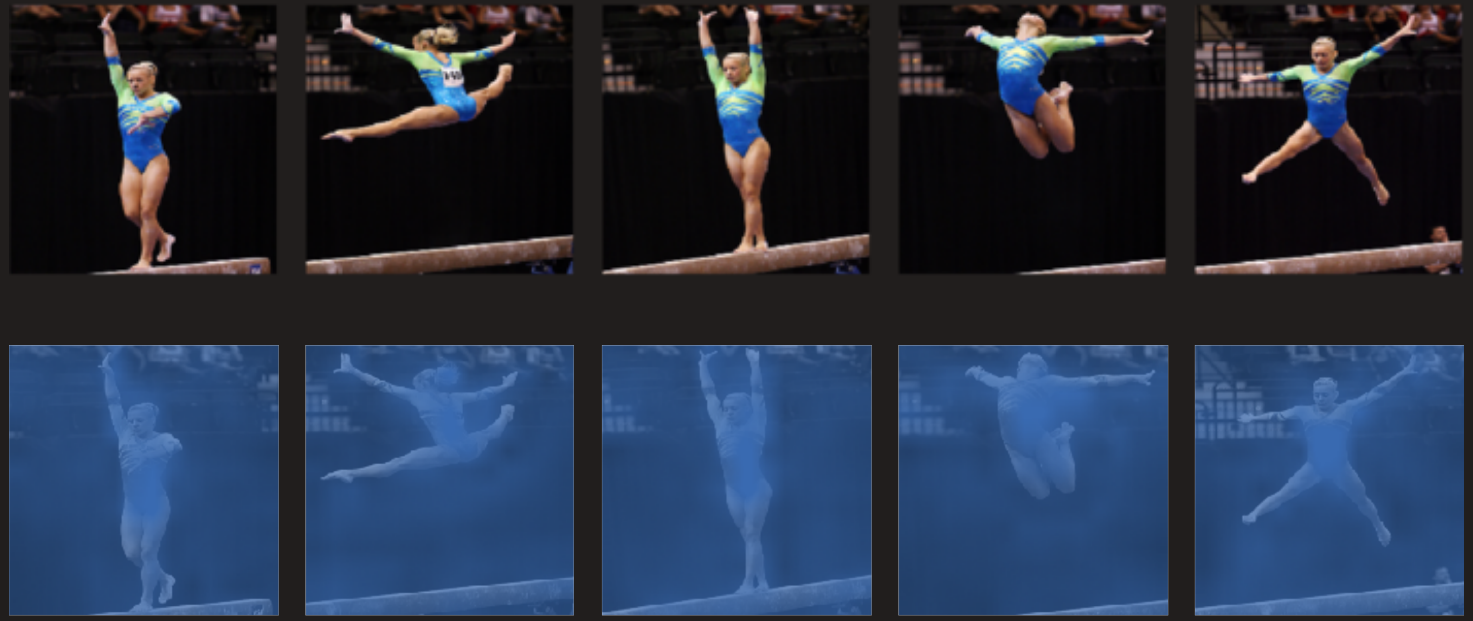
$K = 4$

NMF heatmaps – VGG-19



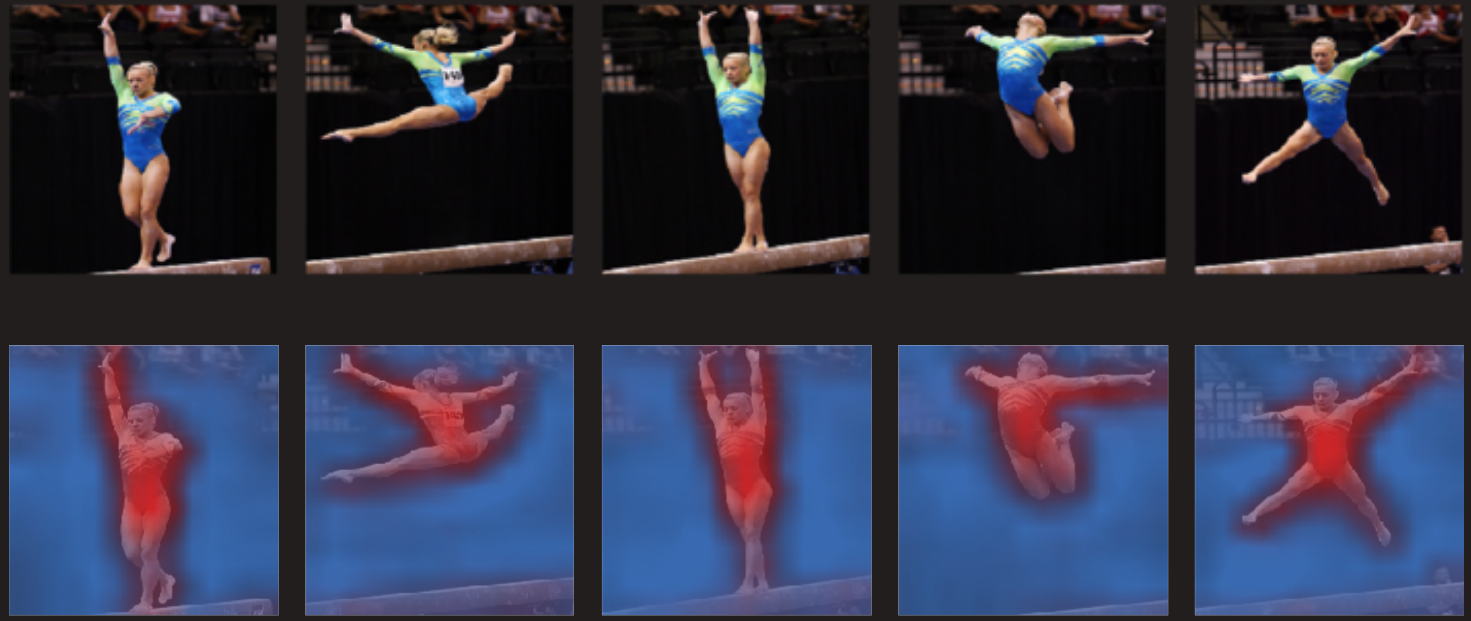
$K = 3$

NMF heatmaps – ResNet-50



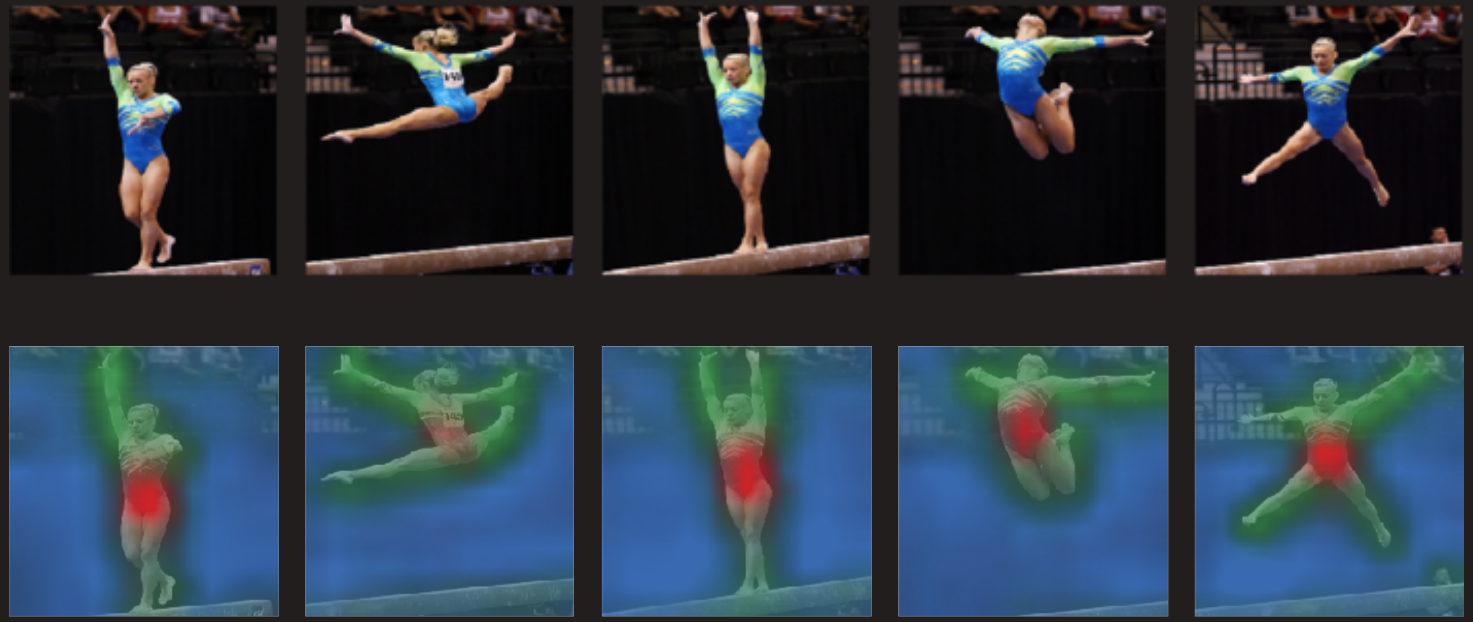
$K = 1$

NMF heatmaps – ResNet-50



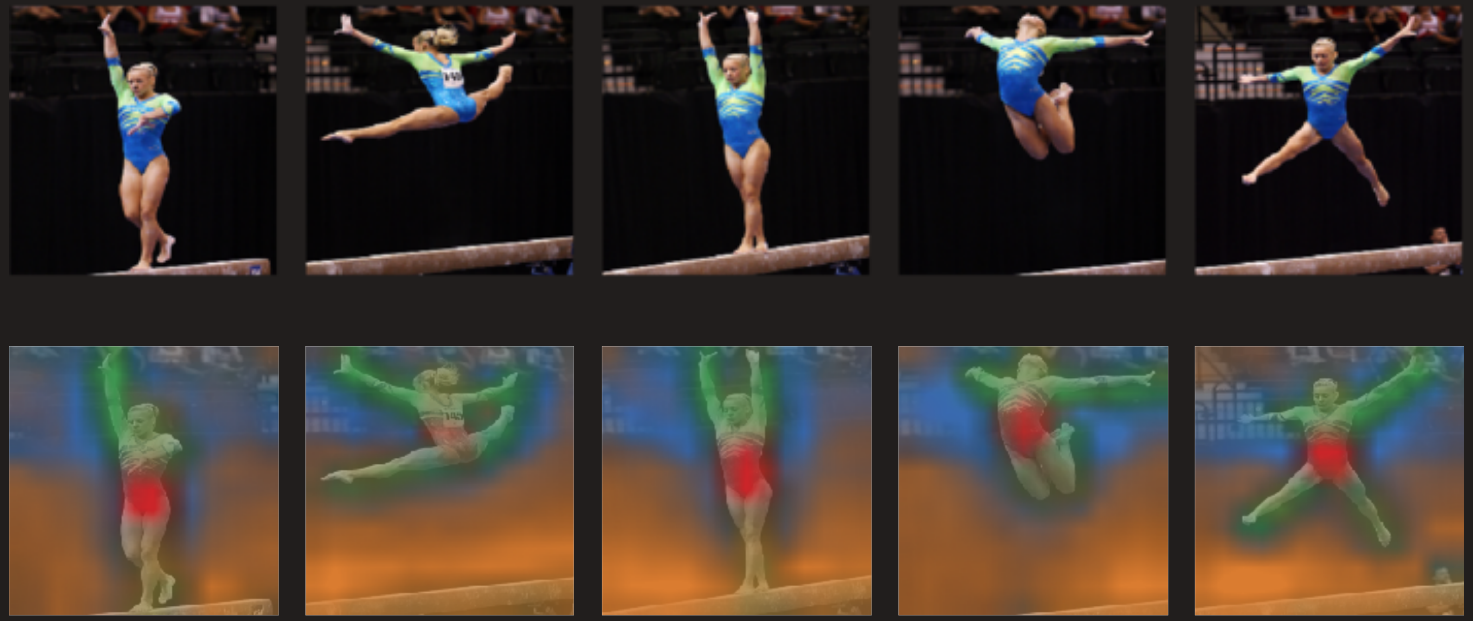
$K = 2$

NMF heatmaps – ResNet-50



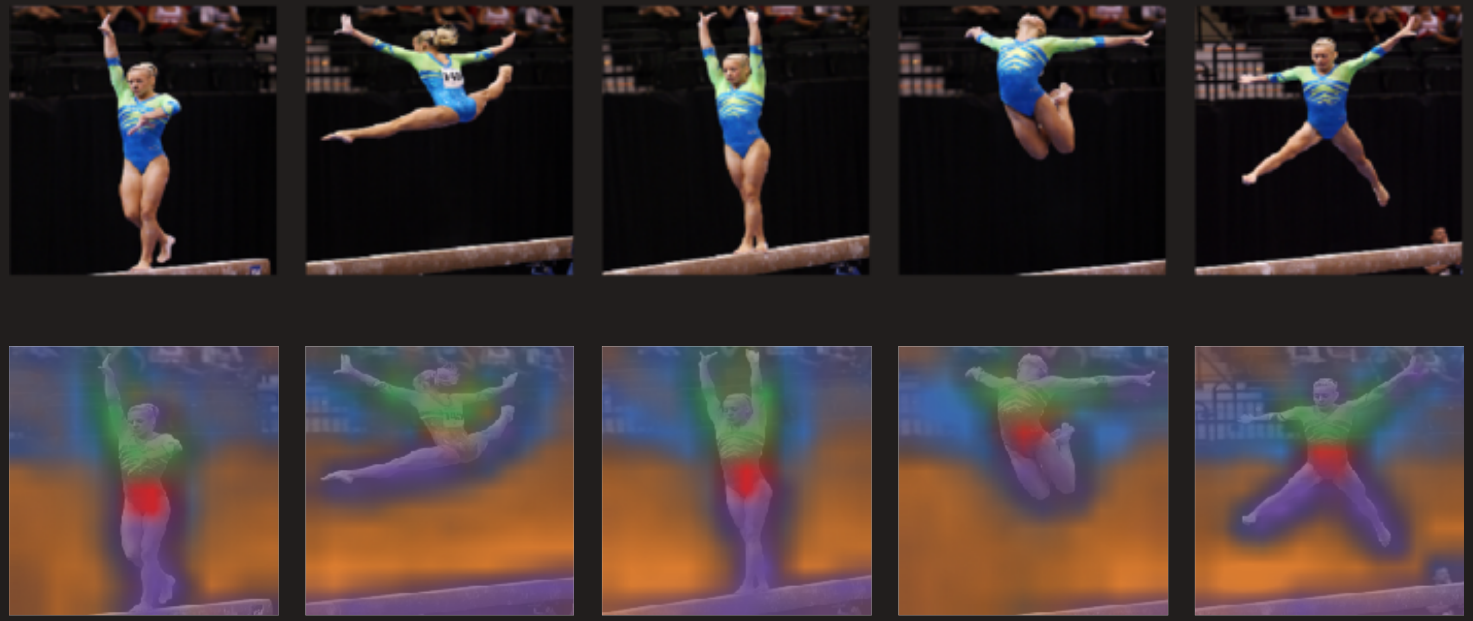
$K = 3$

NMF heatmaps – ResNet-50



$K = 4$

NMF heatmaps – ResNet-50



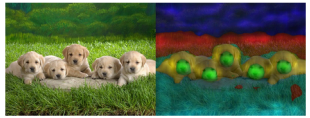
$K = 5$

Semantic localization with matrix U

- The matrix U can be reshaped into K sets of low-resolution heatmaps.
- Each set corresponds to a *semantic part*, based on learned CNN invariance.
- Increasing K reveals a *part hierarchy* in feature space.
- Residual connections (e.g., ResNet) lead to distinctly different matrix factors.
- Co-segmentation/co-localization results as good as or better than domain-specific methods.

Edo Collins, Radhakrishna Achanta, Sabine Süsstrunk, Deep Feature Factorization for Concept Discovery, ECCV 2018

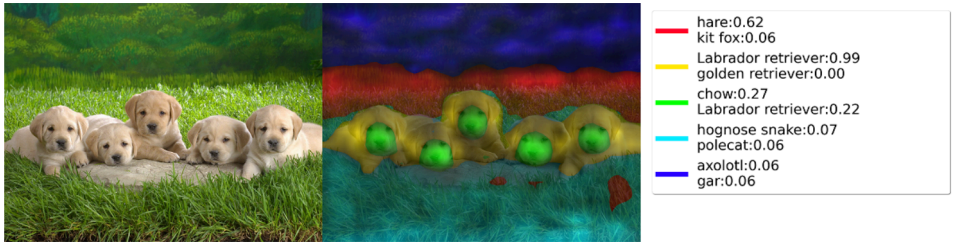
pytorch-gradcam



Advanced AI explainability with pytorch-gradcam

Search this book...

Deep Feature Factorizations for better model explainability



In this tutorial we will see how a method called “Deep Feature Factorizations” can be used for creating insightfull visualizations about what the models see inside images. The pytorch-gradcam package provides an implementation of this method and some additions that make this a very useful tool.

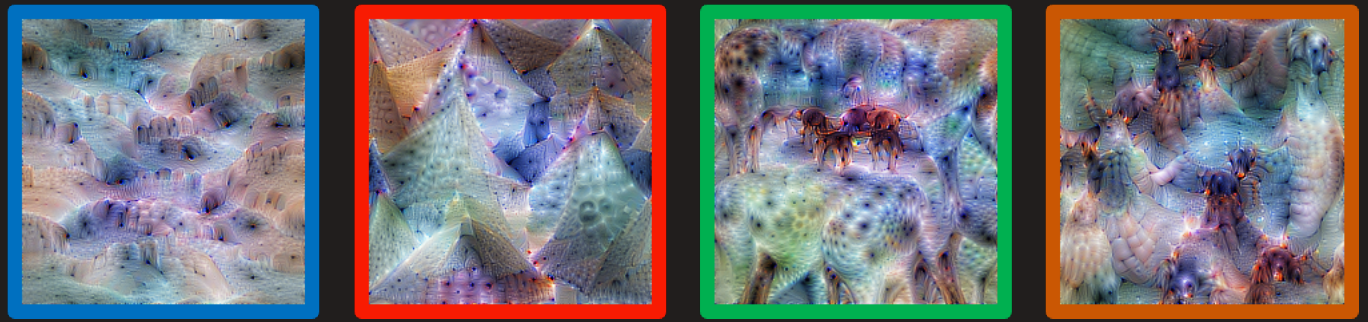
Usually explainability methods answer questions like “Where does the model see a cat in the image” ?

Here instead we will get a much more detailed glimpse into the model, and ask it: “Show me all the different concepts you found inside the image, and how are they classified”.

We will go over the motivations for this, problems with previous methods, and hopefully get a tool that solves these problems.

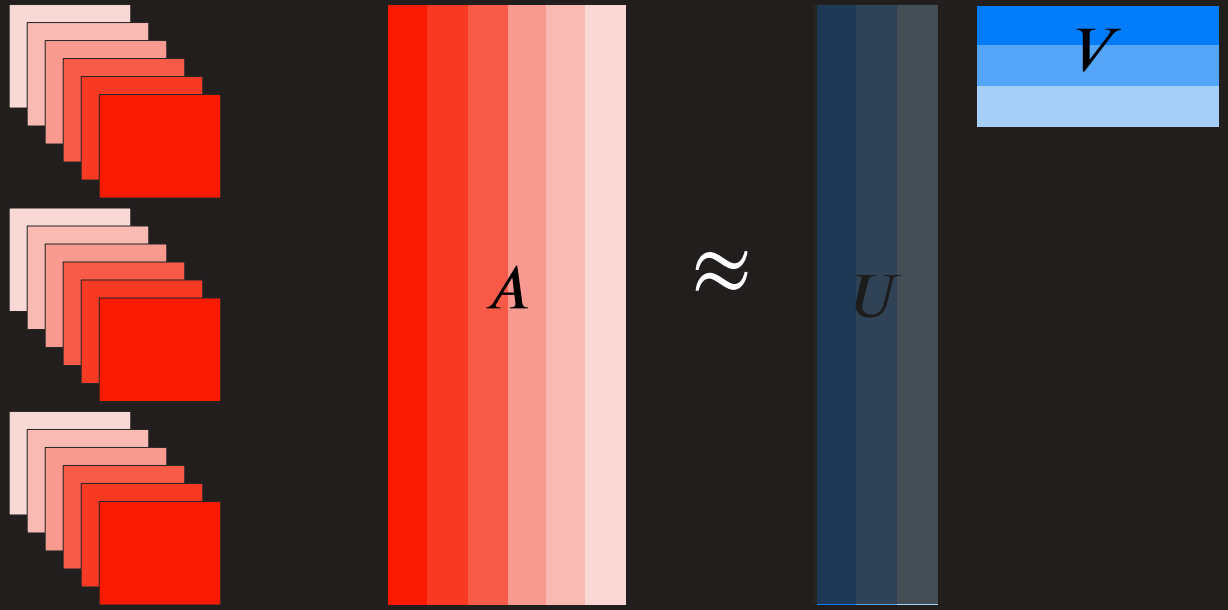
Semantic localization with matrix *U*

Semantic retrieval with matrix *V*



V is a global dictionary

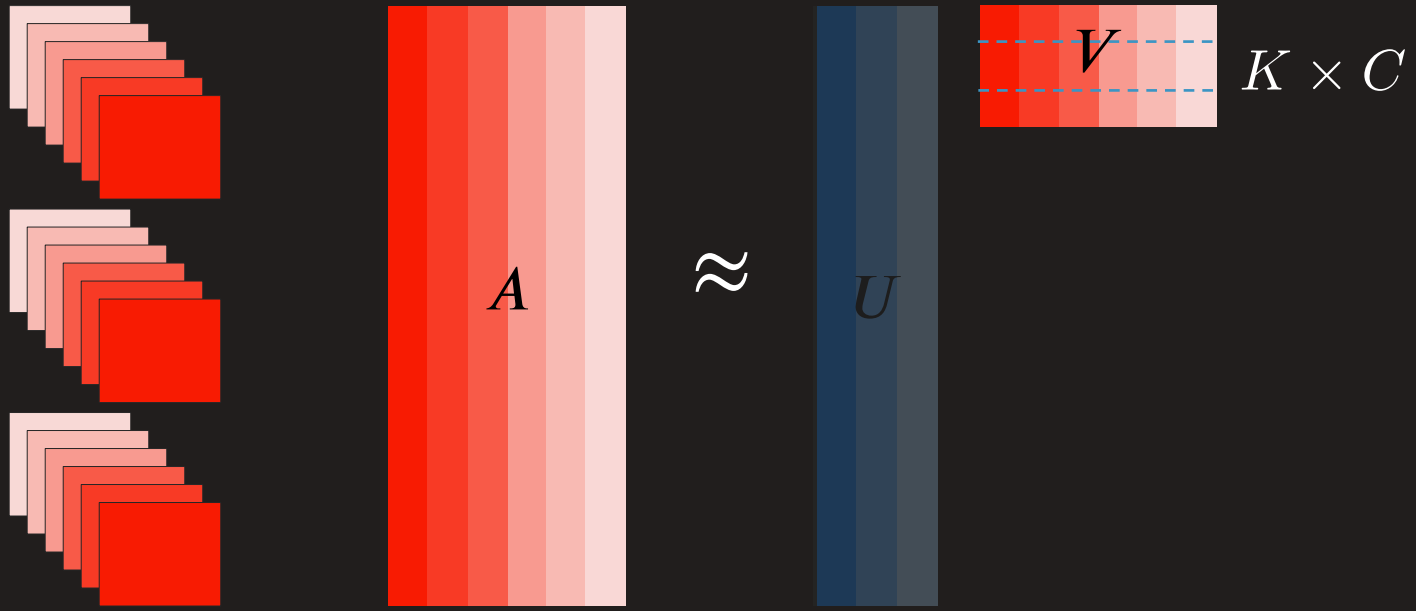
$$N \times C \times H \times W \rightarrow (N \cdot H \cdot W) \times C$$



E. Collins, S. Süsstrunk. *Deep Feature Factorization for Content-based Image Retrieval and Localization*. IEEE ICIP 2019

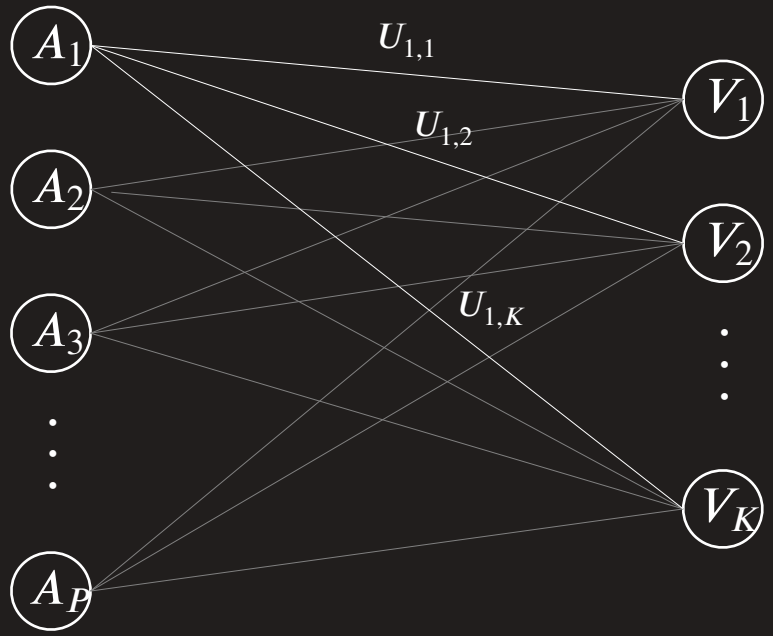
V is a global dictionary

$$N \times C \times H \times W \rightarrow (N \cdot H \cdot W) \times C$$



E. Collins, S. Süsstrunk. *Deep Feature Factorization for Content-based Image Retrieval and Localization*. IEEE ICIP 2019

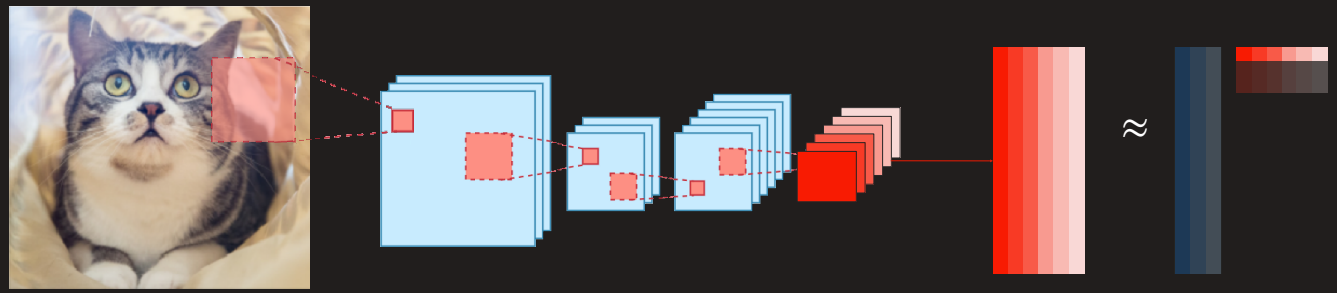
V is a global dictionary



E. Collins, S. Süsstrunk. *Deep Feature Factorization for Content-based Image Retrieval and Localization*. IEEE ICIP 2019

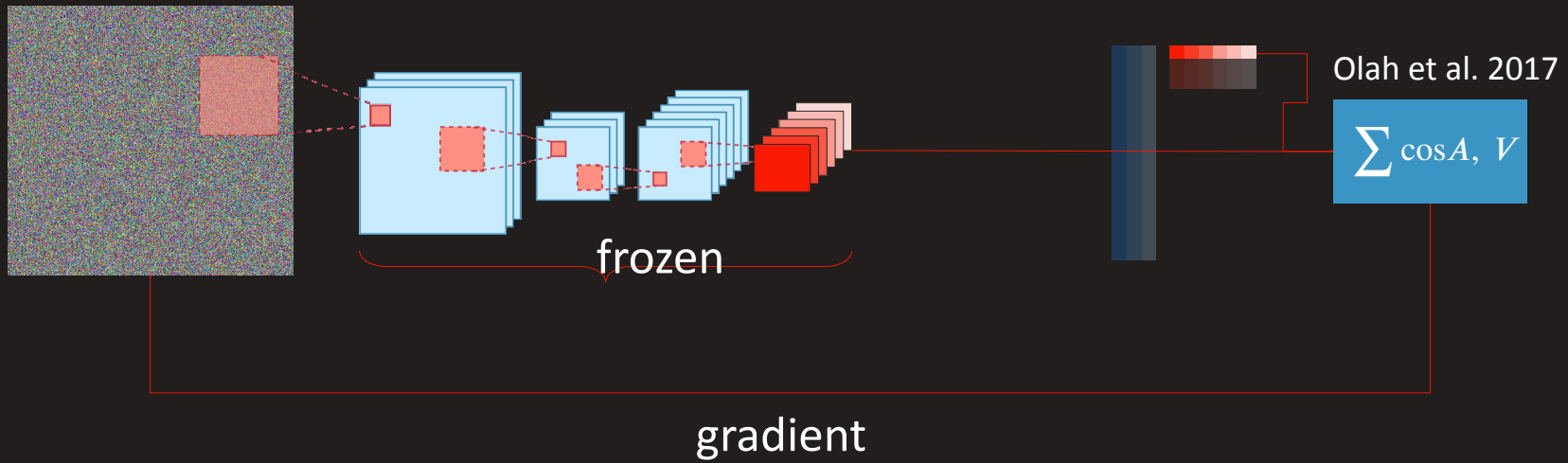
Gradient ascent visualization of V

$$\mathbf{I}_k = \arg\max_{\mathbf{I}} \sum_{i,j} \cos\left(\mathbf{A}_{.,i,j}, \boldsymbol{V}_{k,.}\right)_j, 1 \leq k \leq K$$

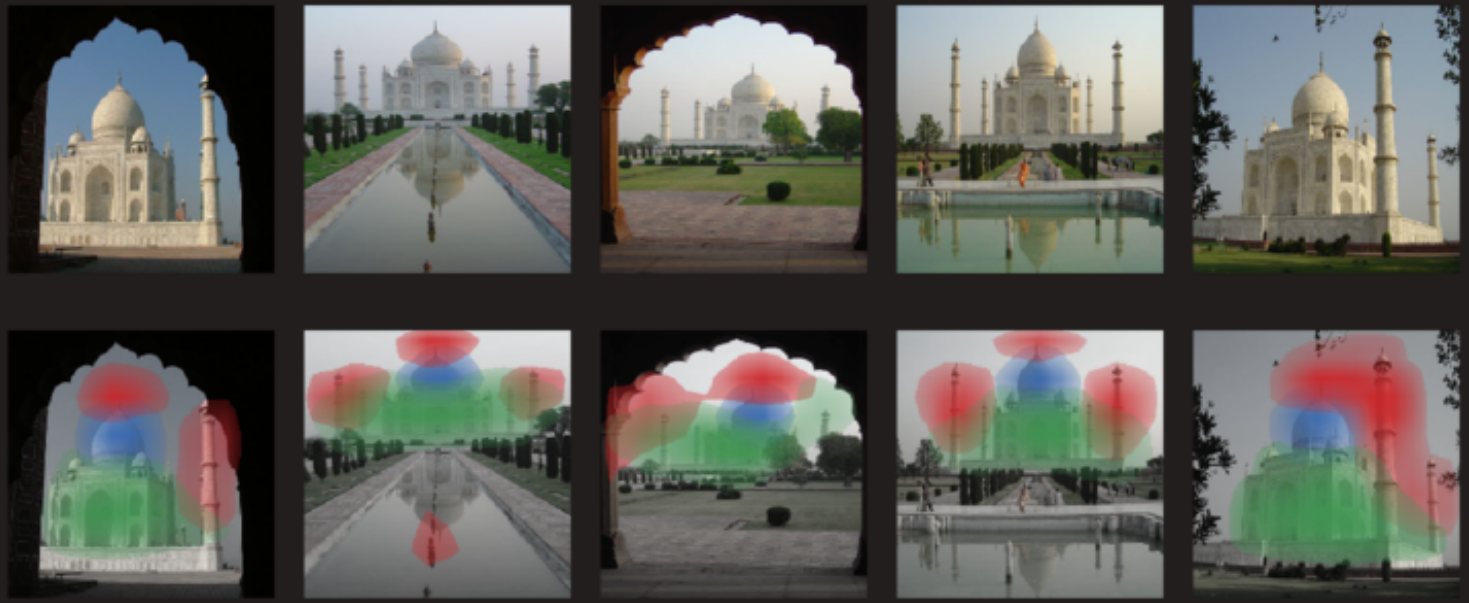


Gradient ascent visualization of V

$$\mathbf{I}_k = \arg\max_{\mathbf{I}} \sum_{i,j} \cos\left(\mathbf{A}_{.,i,j}, \boldsymbol{V}_{k,.}\right)_j, 1 \leq k \leq K$$

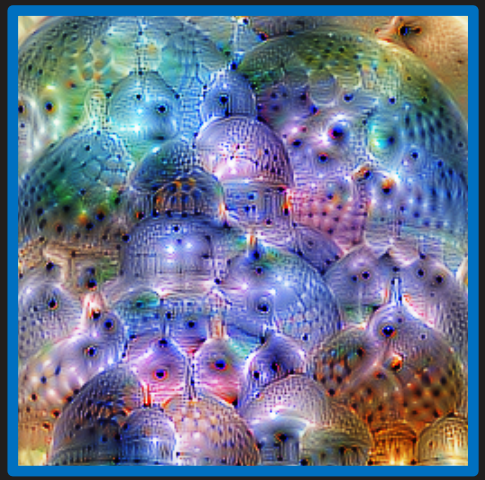


Gradient ascent visualization of V – VGG-19



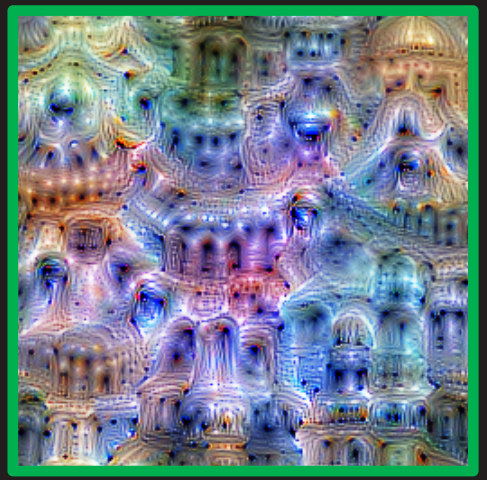
$K = 3$

Gradient ascent visualization of V – VGG-19



Gradient ascent visualization of V – VGG-19

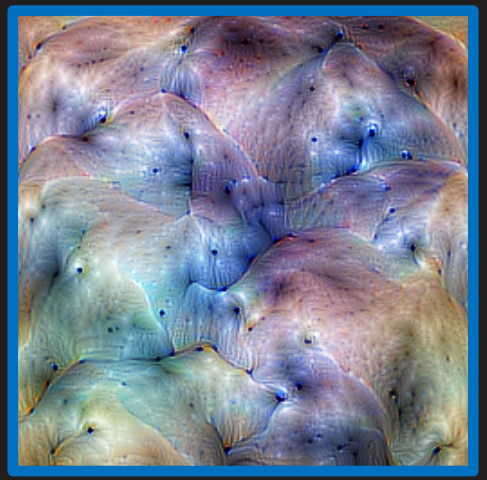
Gradient ascent visualization of V – VGG-19



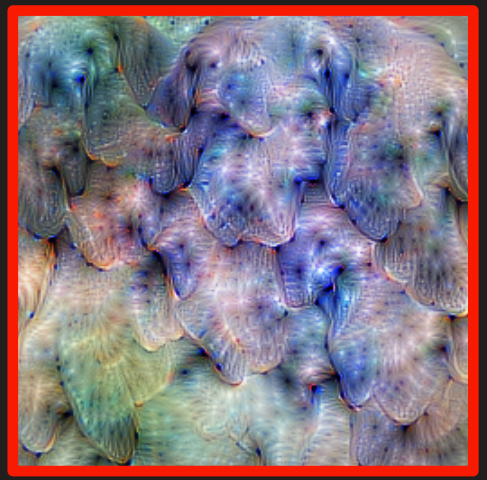
Gradient ascent visualization of V – VGG-19

$K = 4$

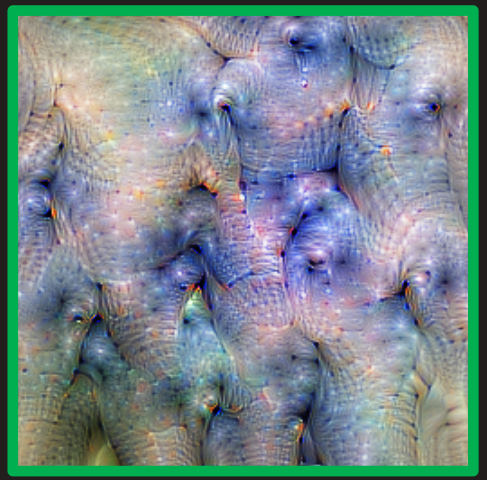
Gradient ascent visualization of V – VGG-19



Gradient ascent visualization of V – VGG-19



Gradient ascent visualization of V – VGG-19

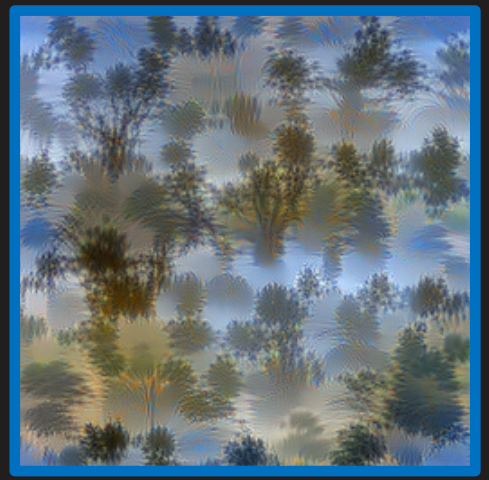


Gradient ascent visualization of V – ResNet50

Gradient ascent visualization of V – ResNet50

$K = 5$

Gradient ascent visualization of V – ResNet50



Gradient ascent visualization of V – ResNet50

Gradient ascent visualization of V – ResNet50

Gradient ascent visualization of V – ResNet50

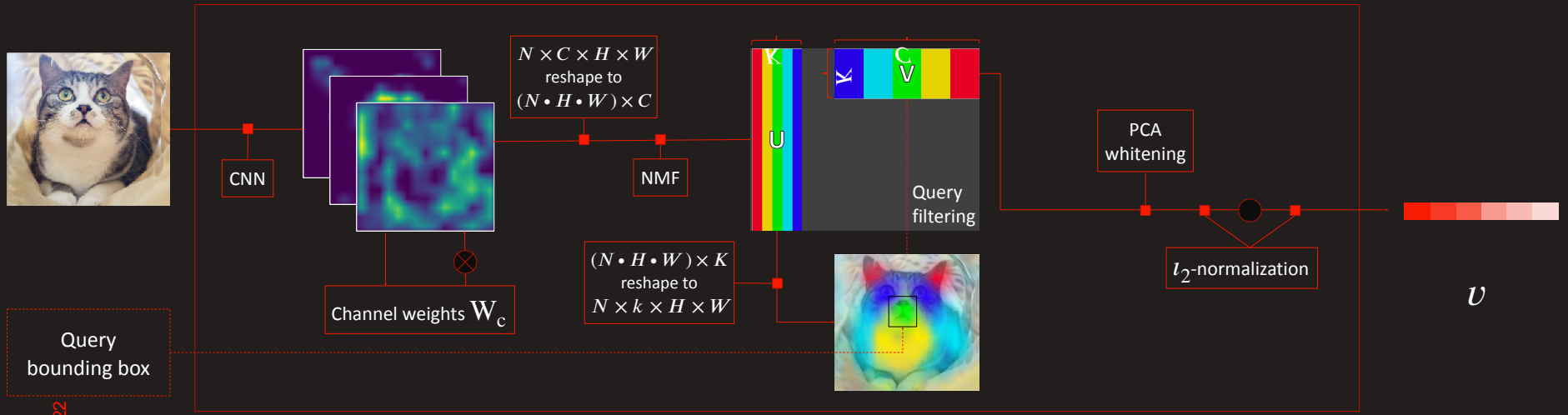
Gradient ascent visualization of V – ResNet50

Content-based image retrieval

- *Instance-based* image retrieval
  - Relevant images are those depicting the same *instance*

- *Semantic* image retrieval
  - Relevant images are those depicting the same *class*

Content-based image retrieval



E. Collins, S. Süsstrunk. *Deep Feature Factorization for Content-based Image Retrieval and Localization*. IEEE ICIP 2019

Semantic image retrieval 1/2

Network	Method	aero	bicy	bird	boa	bot	bus	car	cat	cha	cow
ResNet-50	MAC	13.5	22.4	10.3	6.7	5.6	9.6	14.8	7.7	5.6	4.6
	CroW	32.5	37.3	17.6	12.9	11.2	23.6	36.8	16.3	11.3	7.6
	R-MAC	38.1	48.4	30.8	14.8	15.2	27.0	21.6	26.9	13.2	6.5
	NMF $\boldsymbol{V}+\boldsymbol{U}$	71.3	56.4	46.4	43.5	37.8	57.6	72.7	40.0	25.3	13.3
	NMF $\boldsymbol{V}$	91.9	83.4	54.2	53.1	53.2	72.8	85.4	58.0	31.6	11.9
VGG-16	MAC	43.4	42.9	42.5	24.0	30.8	49.5	46.9	38.9	14.4	11.1
	CroW	76.5	69.0	58.8	38.6	35.6	64.9	64.0	52.6	29.1	22.3
	R-MAC	83.5	82.5	70.3	49.3	49.2	69.4	71.9	57.0	28.1	19.0
	NMF $\boldsymbol{V}+\boldsymbol{U}$	96.4	67.6	64.3	62.9	64.6	80.7	87.9	84.1	29.3	25.2
	NMF $\boldsymbol{V}$	**100**	**97.5**	**83.5**	**82.6**	**72.5**	**92.5**	**95.3**	**90.1**	**44.6**	**31.0**

Mean average precision @ 30

E. Collins, S. Süsstrunk. *Deep Feature Factorization for Content-based Image Retrieval and Localization*. IEEE ICIP 2019

Semantic image retrieval 2/2

Network	Method	dtab	dog	hors	mbik	pers	plnt	she	sofa	trai	tv	Mean
ResNet-50	MAC	3.9	18.1	9.2	9.5	24.8	9.0	10.1	3.5	6.9	13.9	10.5
	CroW	10.5	27.6	19.2	25.0	43.3	16.9	8.6	10.0	7.7	21.6	19.9
	R-MAC	7.2	33.2	28.1	24.7	38.4	23.3	21.4	10.3	11.6	22.9	23.2
	NMF $\boldsymbol{V}+\boldsymbol{U}$	20.0	43.9	39.9	33.3	60.0	30.5	31.5	32.7	34.2	41.6	41.6
	NMF $\boldsymbol{V}$	26.0	53.7	47.2	60.3	89.1	34.7	38.2	42.3	35.8	43.5	53.3
VGG-16	MAC	5.9	57.6	22.7	22.9	48.1	23.8	31.0	11.3	24.4	29.3	31.1
	CroW	17.1	64.8	54.3	56.4	71.1	33.6	49.1	41.2	37.8	54.9	49.6
	R-MAC	14.6	59.9	53.6	54.7	74.7	55.2	62.9	39.5	53.7	64.6	55.7
	NMF $\boldsymbol{V}+\boldsymbol{U}$	28.6	74.4	71.8	67.4	76.7	52.8	70.7	43.8	42.8	81.5	63.7
	NMF $\boldsymbol{V}$	**29.4**	**86.0**	**82.6**	**94.2**	**81.2**	**71.9**	**86.8**	**51.1**	**56.0**	**95.8**	**76.2**

Mean average precision @ 30

E. Collins, S. Süsstrunk. *Deep Feature Factorization for Content-based Image Retrieval and Localization*. IEEE ICIP 2019

Semantic retrieval with matrix V

- V is a global dictionary representing important concepts
- Rows of V can be directly visualized using gradient ascent
- Image descriptors derived from V are useful for semantic image retrieval
- Adding spatial information yields descriptors useful for instance-based image retrieval

What about GANs?

Editing in Style: Uncovering the Local Semantics of GANs

Edo Collins[1] Raja Bala[2] Bob Price[2] Sabine Süsstrunk[1]

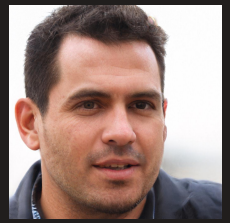
[1] School of Computer and Communication Sciences, EPFL, Switzerland

[2] Interactive and Analytics Lab, Palo Alto Research Center, Palo Alto, CA

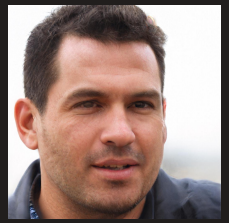
Video

Changing local semantics in an image

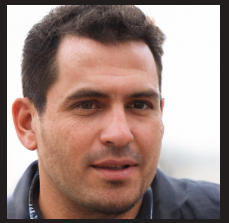
Reference —*style*→ Target



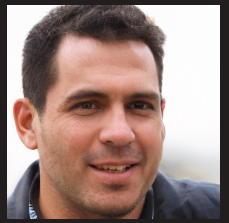
Conditioned on



eyes



nose



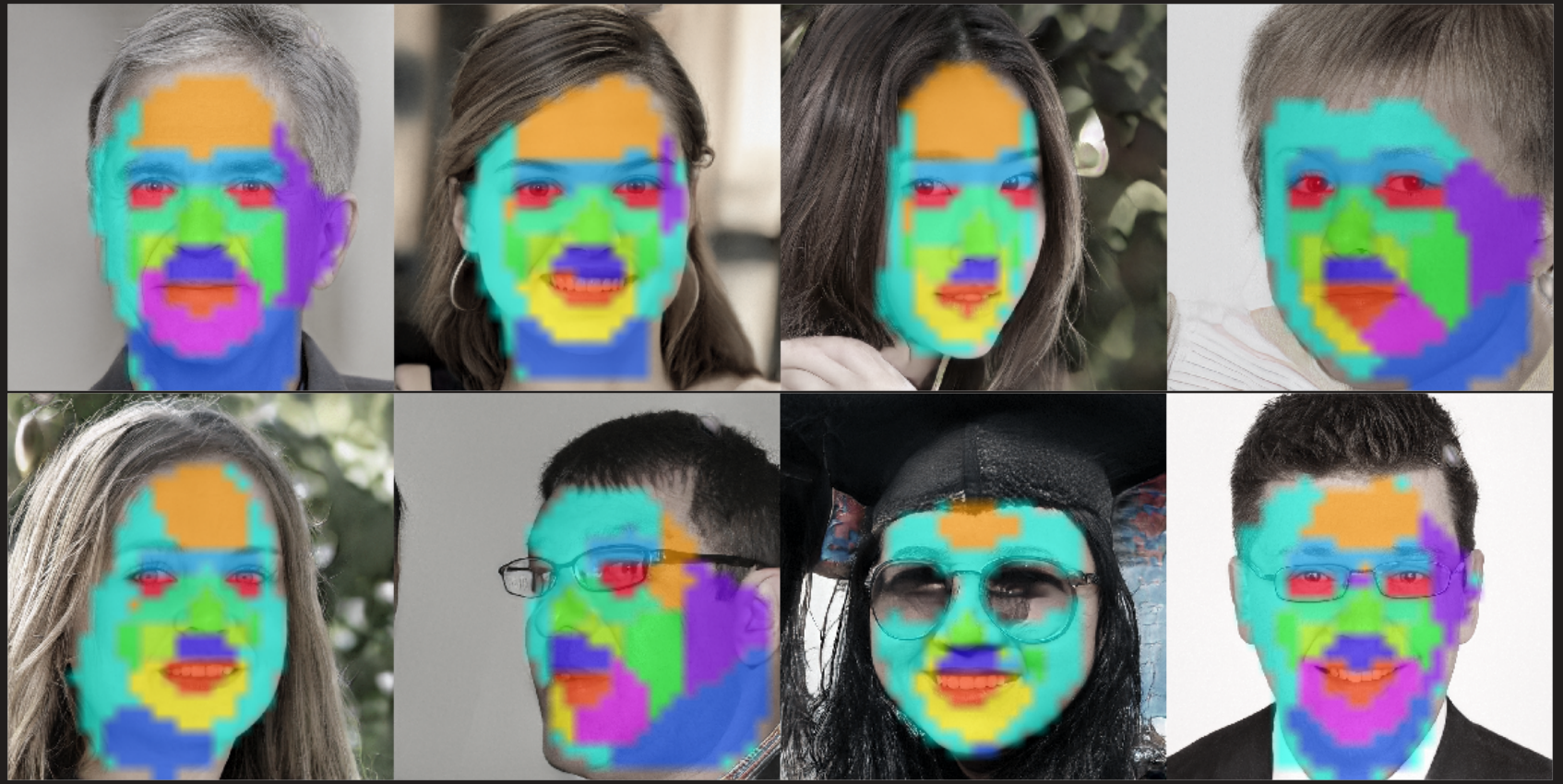
mouth

FFHQ StyleGAN, layer 32x32, K=2

FFHQ StyleGAN, layer 32x32, $K=3$

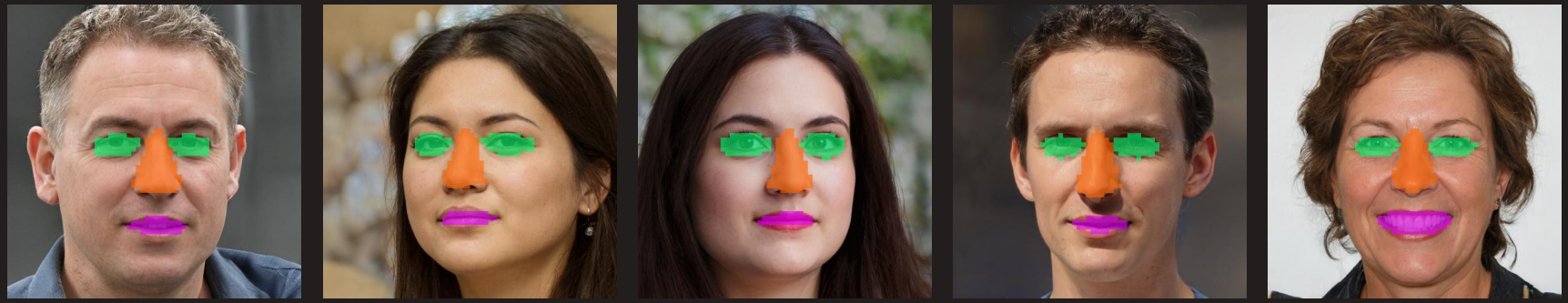
FFHQ StyleGAN, layer 32x32, $K=10$

FFHQ StyleGAN, layer 32x32, K=25



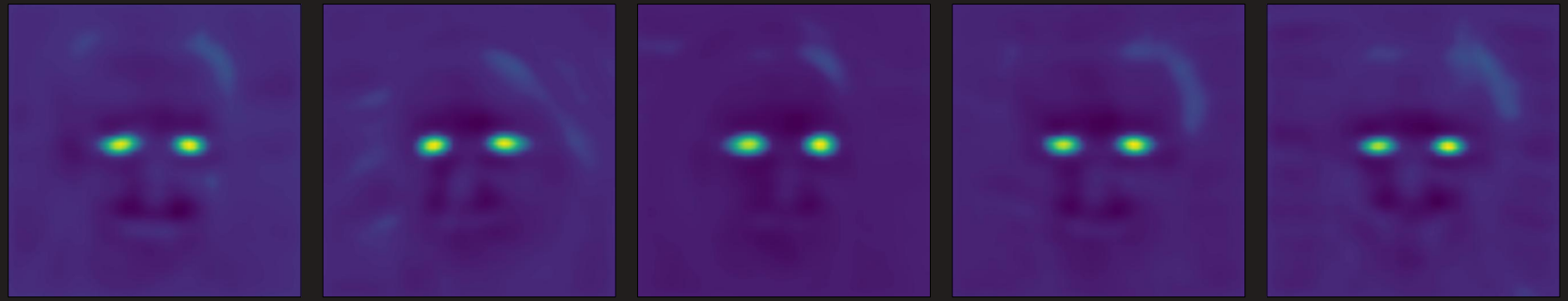
FFHQ StyleGAN, layer 128x128, K=25

Labeling clusters

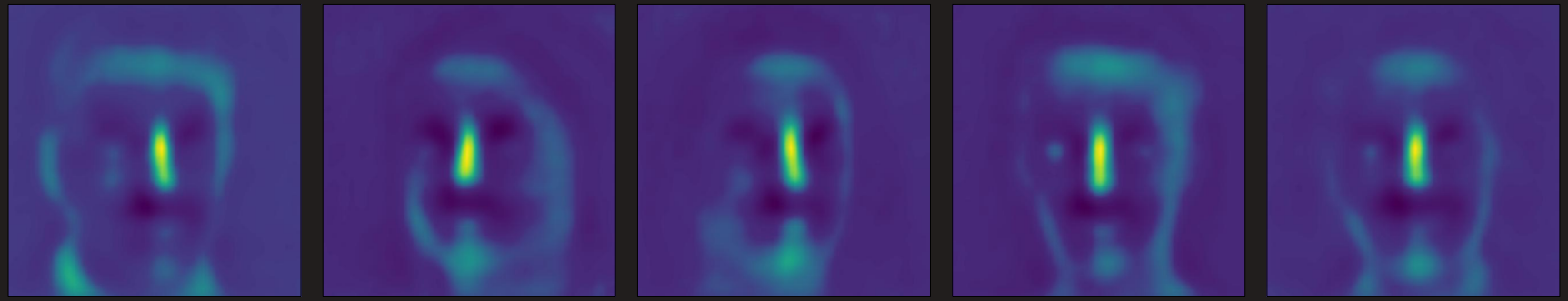


eyes *nose* *mouth*

The best overlapping channel with *eyes*



The best overlapping channel with *nose*



The best overlapping channel with *mouth*

Reference

Our method supports *localized* editing of semantic objects

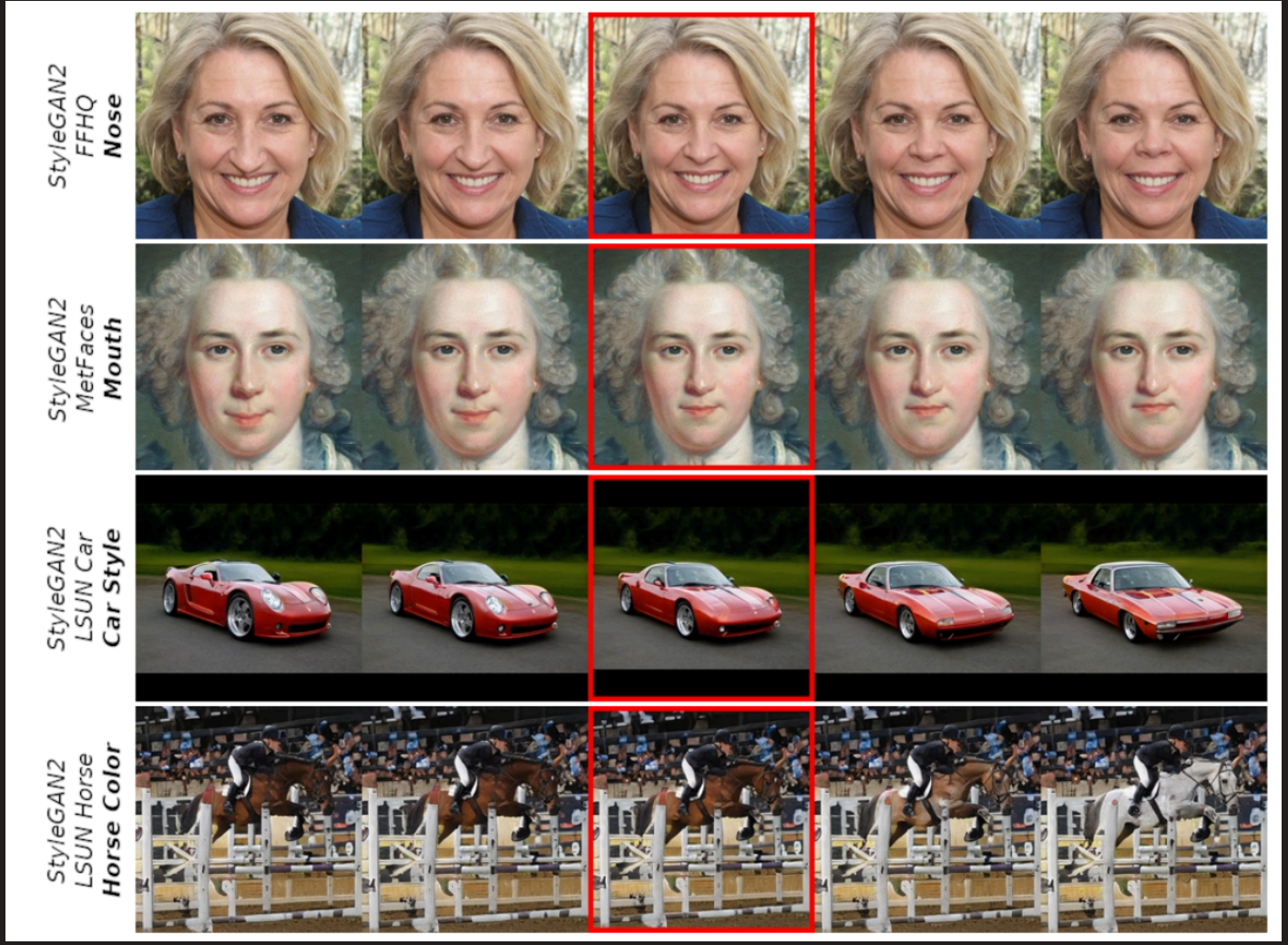
Target

eyes

nose

mouth

Localized edits



E. Pajouheshgar, T Zhang, S Süsstrunk, Optimizing Latent Space Directions for Gan-Based Local Image Editing, ICASSP 2022.

E Pajouheshgar, T Zhang, S Süsstrunk, Optimizing Latent Space Directions for Gan-Based Local Image Editing, ICASSP 2022.

Image Editing with GANs

Original

E. Pajouheshgar, T Zhang, S Süsstrunk, Optimizing Latent Space Directions for Gan-Based Local Image Editing, ICASSP 2022.

Image Editing with GANs

Original

+ hair

E. Pajouheshgar, T Zhang, S Süsstrunk, Optimizing Latent Space Directions for Gan-Based Local Image Editing, ICASSP 2022.

Image Editing with GANs

Original

+ hair

+ Eye movement

E. Pajouheshgar, T Zhang, S Süsstrunk, Optimizing Latent Space Directions for Gan-Based Local Image Editing, ICASSP 2022.

Image Editing with GANs

Original

+ hair

+ Eye movement

+ Lipstick

E. Pajouheshgar, T Zhang, S Süsstrunk, Optimizing Latent Space Directions for Gan-Based Local Image Editing, ICASSP 2022.

Original

+ Hair

+ Eye movement

+ Lipstick

E. Pajouheshgar, T Zhang, S Süsstrunk, Optimizing Latent Space Directions for Gan-Based Local Image Editing, ICASSP 2022.

Conclusions

- Comprehensive neural architecture explainability and interpretability is still a ways off...
- “Old” tools from mathematics, image and video processing are helpful in solving some of the mysteries...
- ...and they can lead to applications to even further improve the photographic experience. experience.

Thank you!

Edo Collins

Ehsan Pajouheshgar