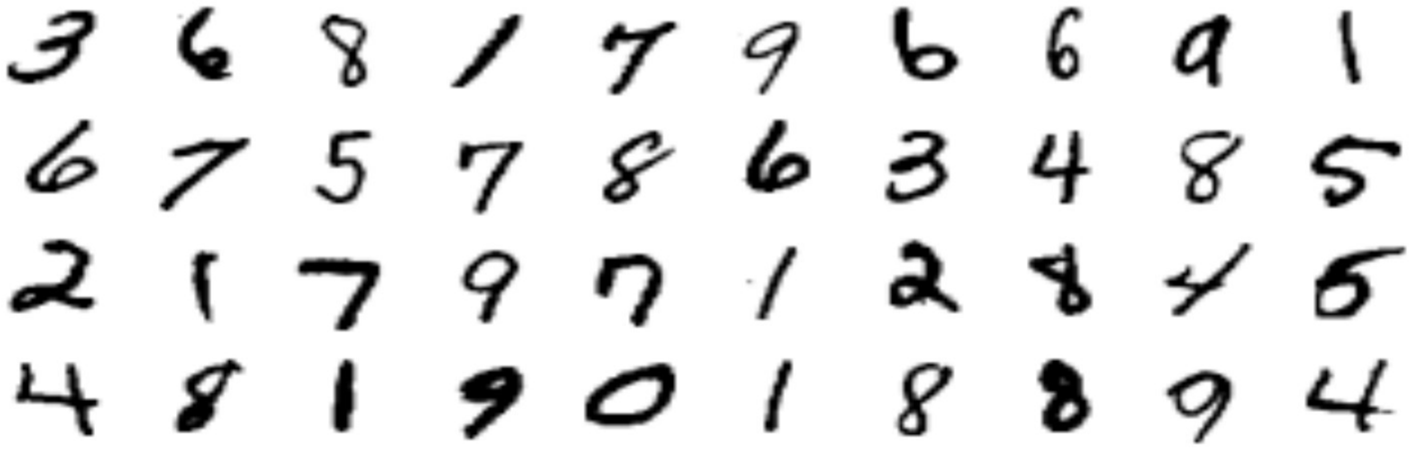
Image classification with patches neighborhood encoding

https://openreview.net/forum?id=aYuZO9DIdnn

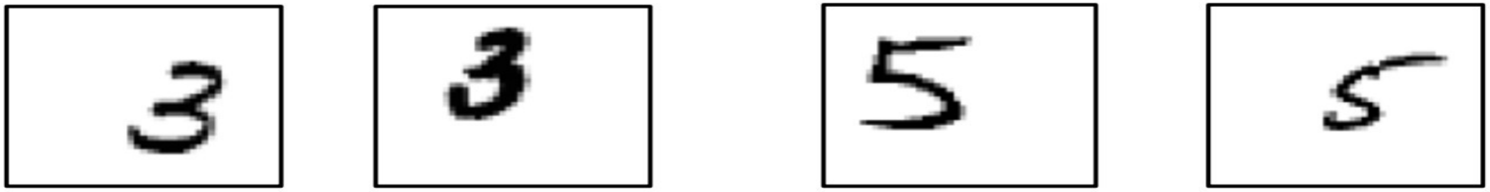
L. Thiry, M. Arbel, E. Belilovsky, E. Oyallon

Digits classification

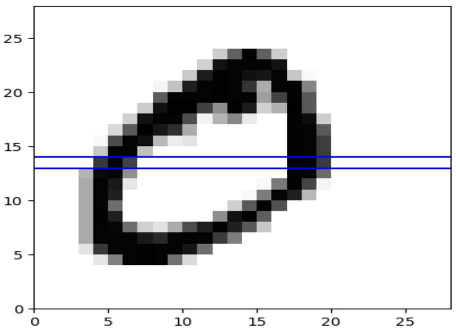
MNIST database

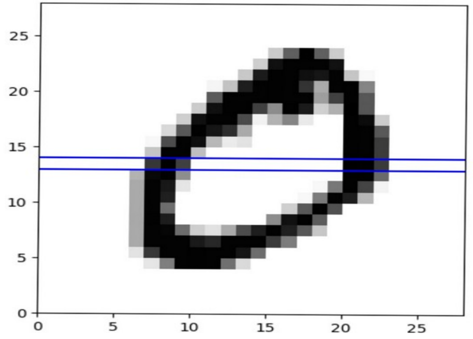


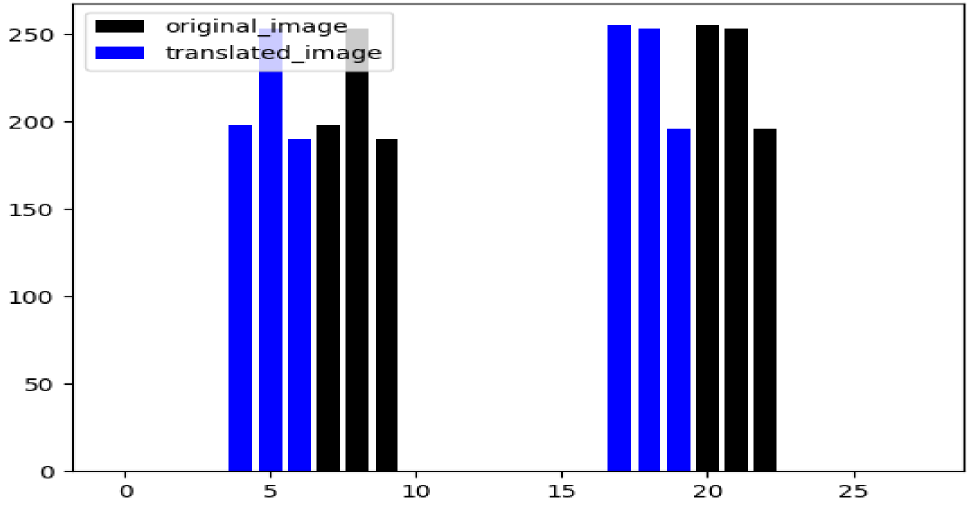
- Invariance to translations, stability to deformations



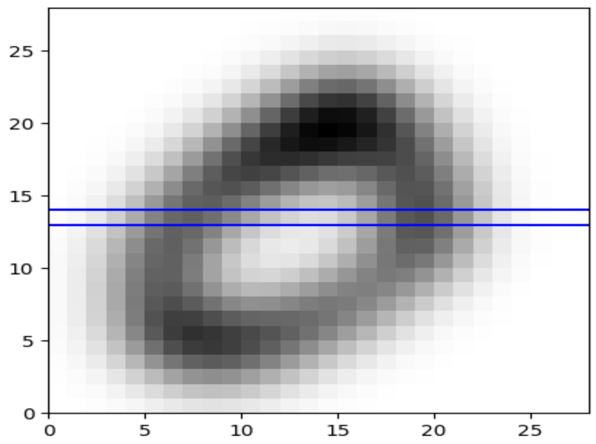
l_2 metric Instability to translations

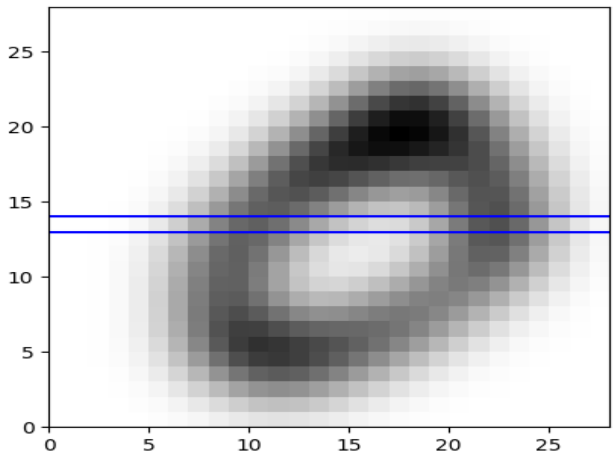


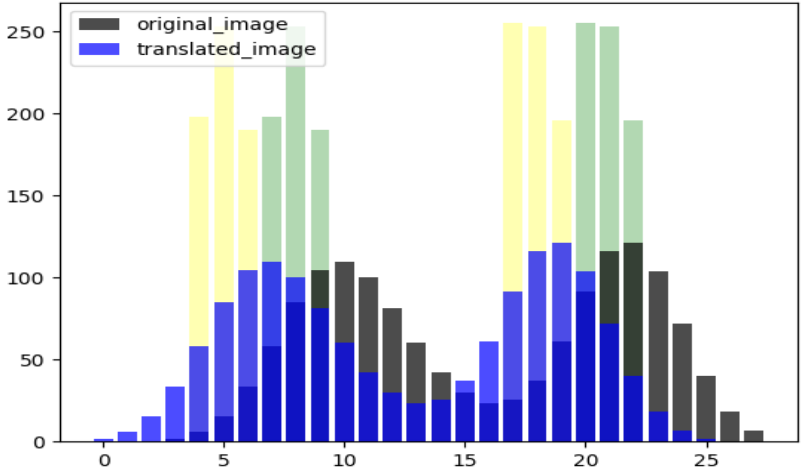




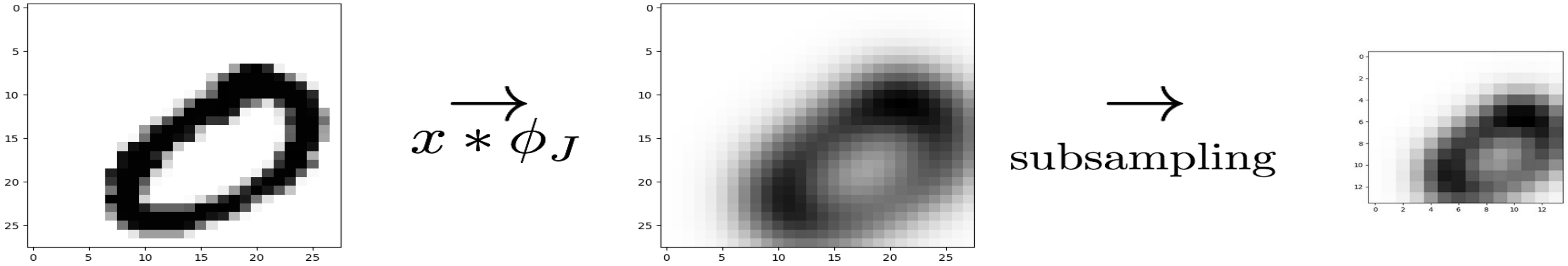
Local averaging







Stability to geometric transformations

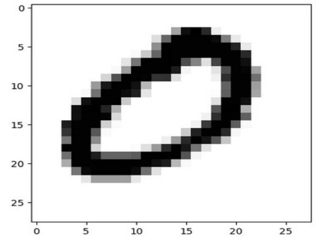


Convolution with Gaussian kernel ϕ_J :

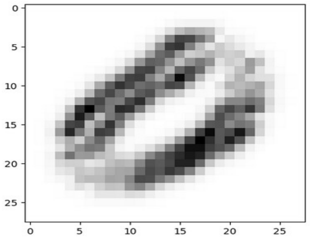
- stable to geometric deformations
- dimensionality reduction via subsampling
- lots of details are lost

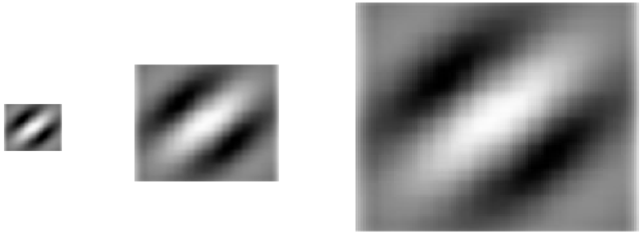
Preserving signal information

Recover information lost in averaging



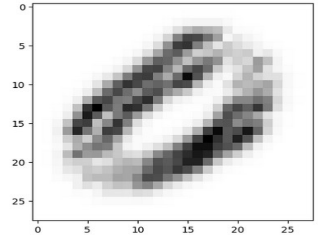
$$\longrightarrow$$
$$|x * \psi_{j,\theta}|$$



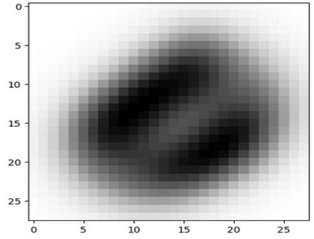


Gabor wavelets $\psi_{j,\theta}$

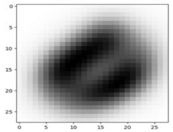
Stability to geometric transformations



$$\longrightarrow$$
$$|x * \psi_{j,\theta}| * \phi_J$$

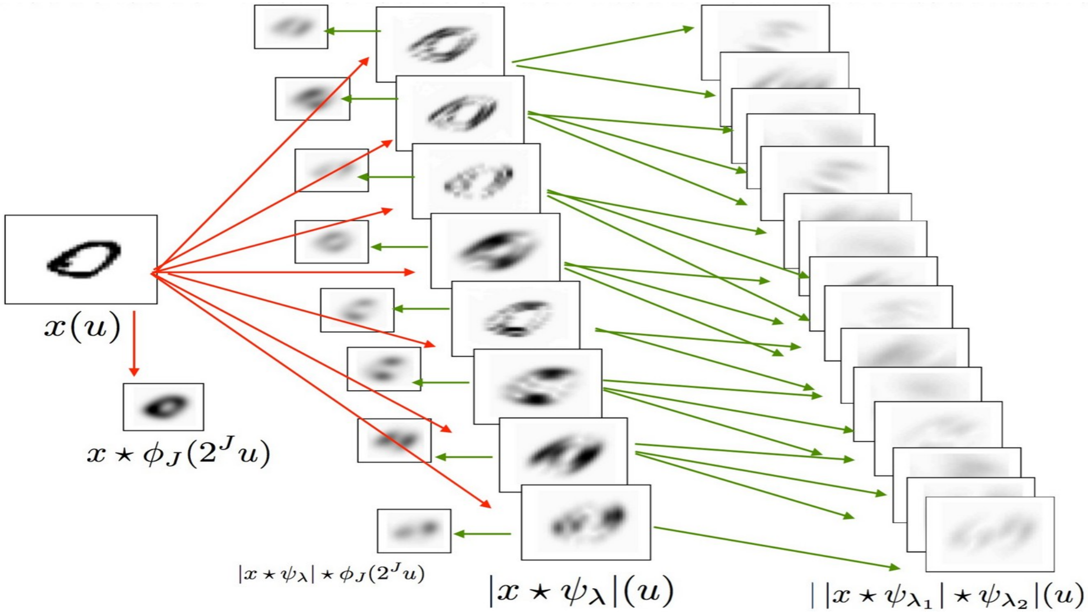


$$\longrightarrow$$
subsampling



Scattering transform

Mallat (2011), Mallat, Bruna (2012)

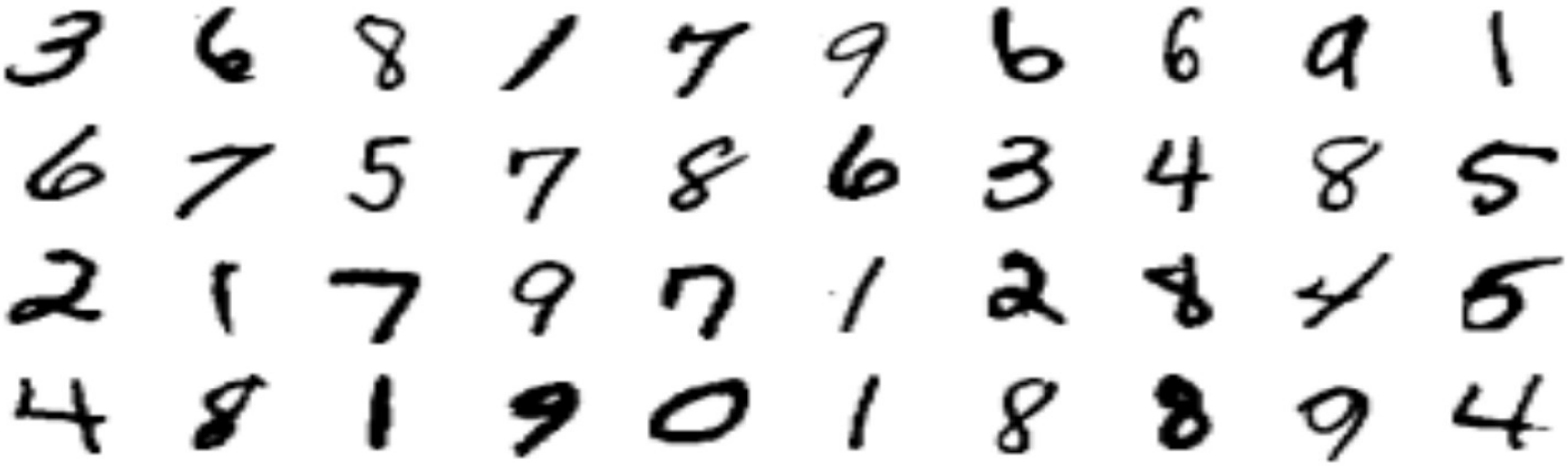


<u>Theorem</u>

$$\|Sx_\tau - Sx\| \leq K\|x\|\|\nabla\tau\|_\infty$$

Scattering vs Deep ConvNets

Dataset	Scattering Transform	AlexNet	ResNet
MNIST 28^2 digit images 10 classes	>99 %	>99 %	>99 %



Scattering vs Deep ConvNets

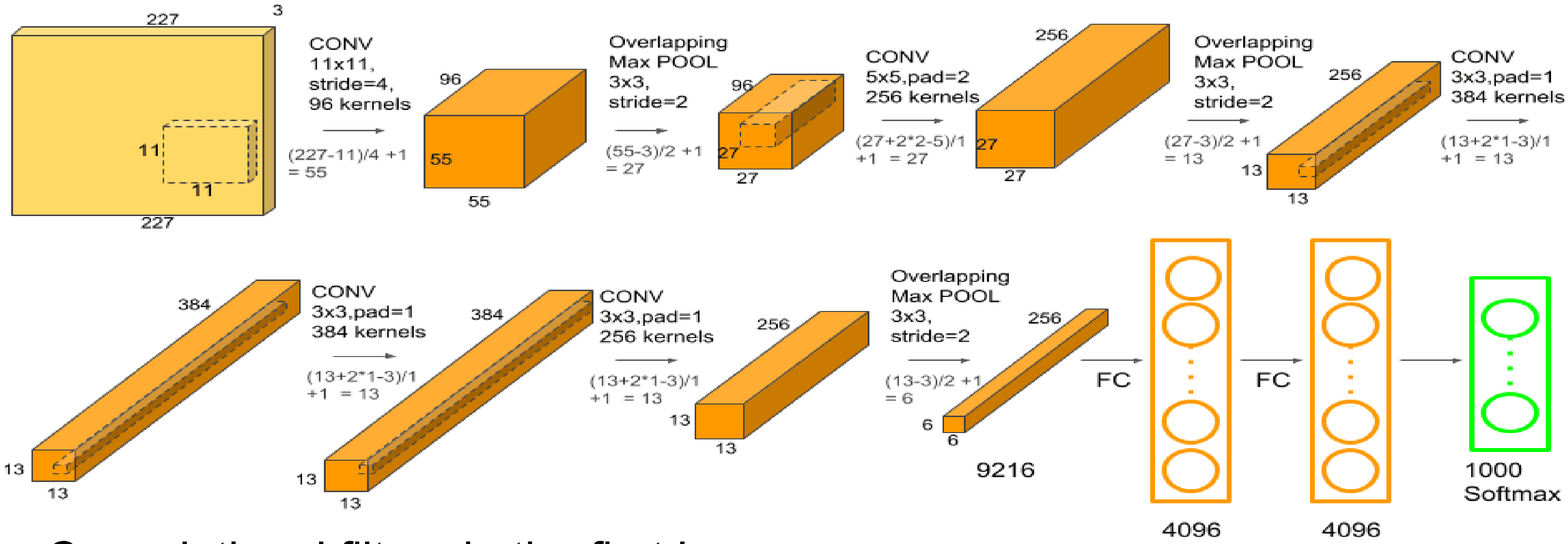
Dataset	Scattering Transform	AlexNet	ResNet
CIFAR-10 32^2 object images 10 classes	82.3 %	89.1 %	95.5 %

Scattering vs Deep ConvNets

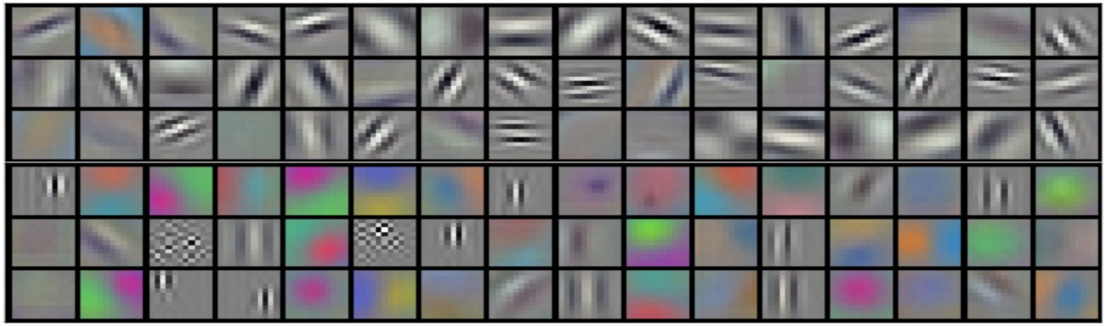
Dataset	Scattering Transform	AlexNet	ResNet
ImageNet 224^2 object images 1000 classes	44.7 %	79.1 %	94.2 %

AlexNet

Krizhevsky et al. 2012



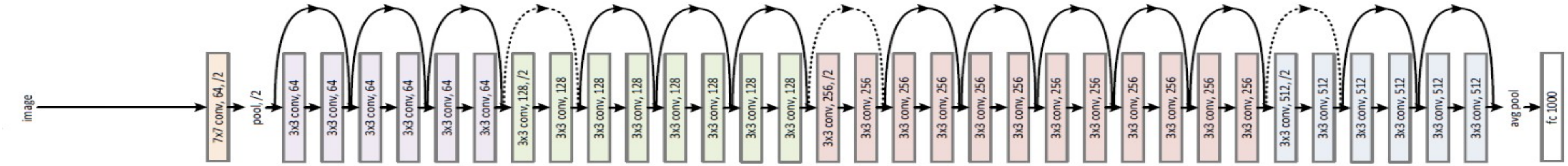
<u>Convolutional filters in the first layer</u>



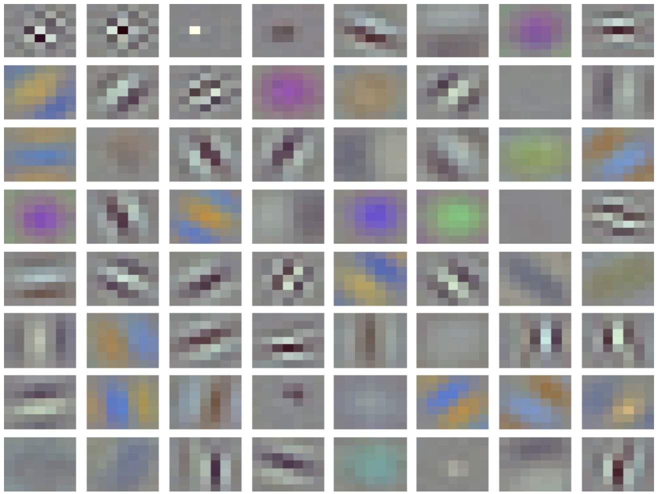
ResNet

He et al. 2016
94.2 % top5 accuracy

- skip connections
- up to 152 convolutional layers



<u>Convolutional filters in the first layer</u>

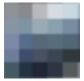


Patch K nearest neighbors binary encoding

Ours

Image patch

Small square region of the image



Mahalanobis distance

random vector X with covariance $\Sigma = P\Lambda P^T$

$$D_M(x, x') = \sqrt{(x - x')^T \Sigma^{-1} (x - x')}$$

whitening operator w

$$\mathrm{Cov}(w(\mathbf{X})) = I_n$$

$$w : \mathbf{X} \mapsto O\Lambda^{-1/2} P^T (\mathbf{X} - \mu), \quad \forall O \in O_n(\mathbb{R})$$

$$\|w(x) - w(x')\| = D_M(x, x')$$

Patch K nearest neighbors binary encoding

Ours

Method

- Randomly select a set D of patches
- Regularized whitening operator $W = (\lambda I + \Sigma)^{-1/2}$
- For each image patch $p_{i,x}$ compute set of Mahanalobis distances

$$\mathcal{C}_{i,x} = \{\|W p_{i,x} - W d\| \; d \in \mathcal{D}\}$$

- K nearest neighbors encoding

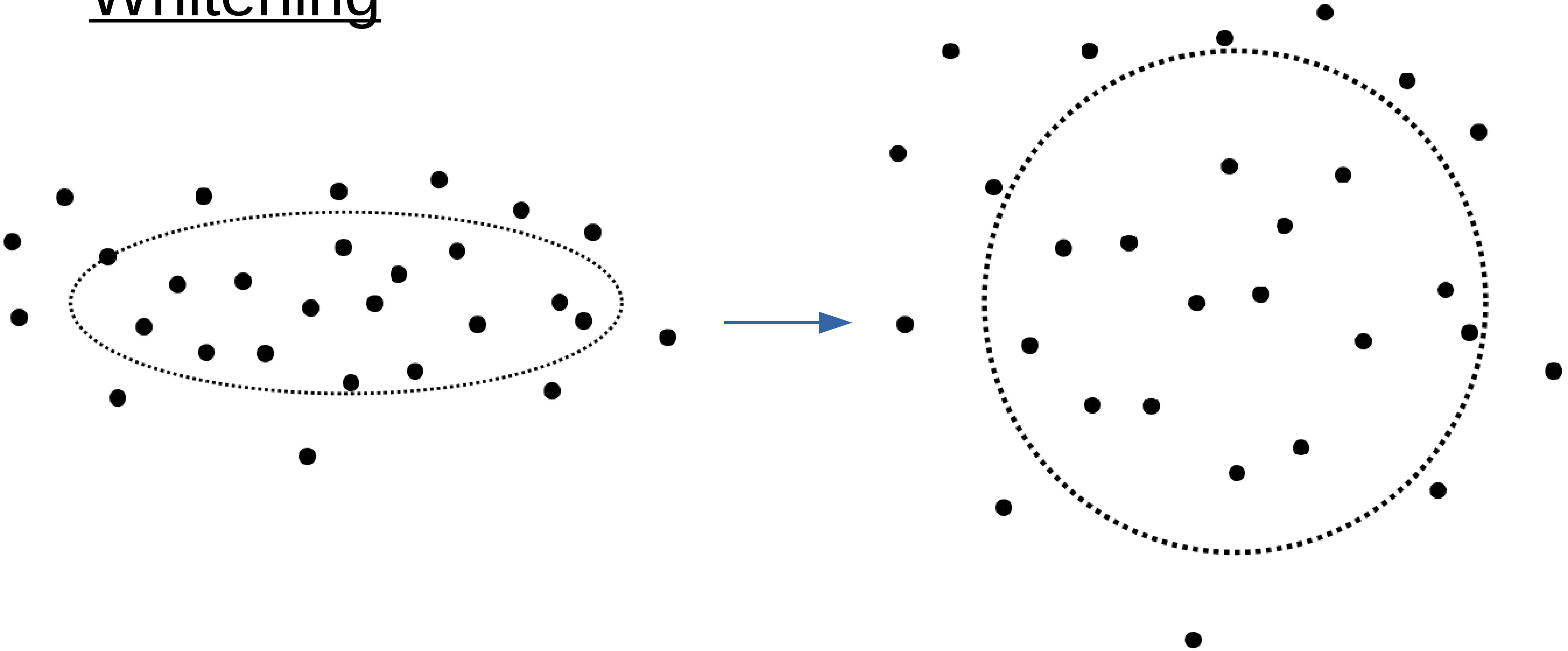
$\tau_{i,x}$ the K-th smallest element of $\mathcal{C}_{i,x}$

$$\phi(x)_{d,i} = \begin{cases} 1, & \text{if } \|p_{i,x} - d\| \leq \tau_{i,x} \\ 0, & \text{otherwise.} \end{cases}$$

Patch K nearest neighbors binary encoding

Ours

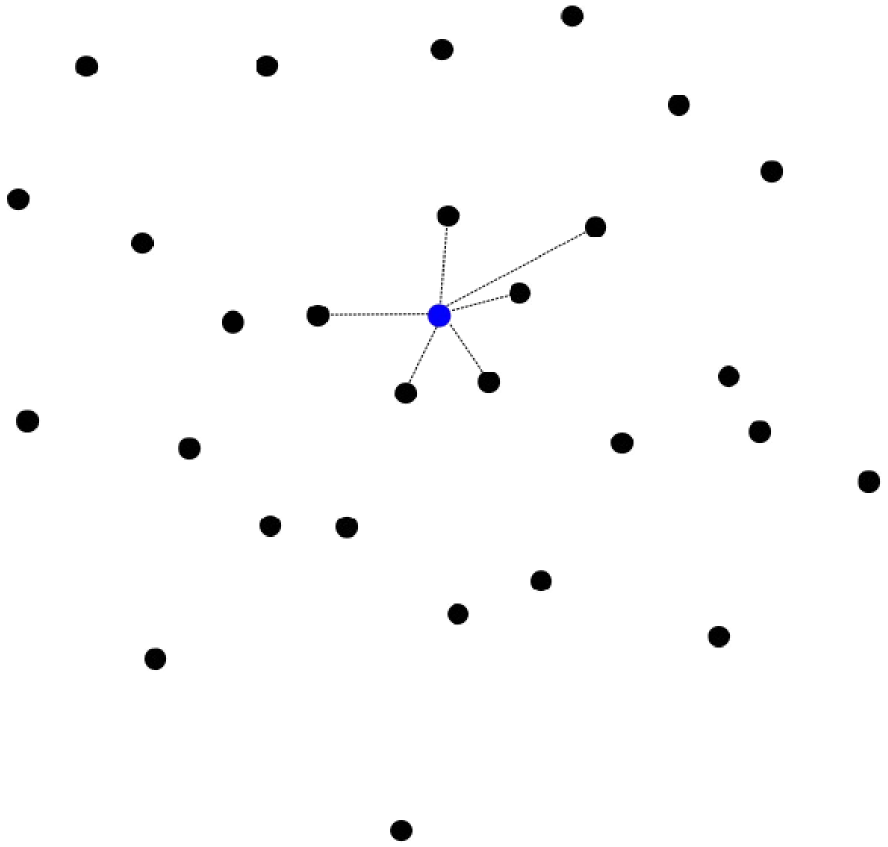
Whitening



Patch K nearest neighbors binary encoding

Ours

K nearest neighbors



Patch K nearest neighbors binary encoding

Ours

Linear classification on CIFAR-10

Method	$\lvert\mathcal{D}\rvert$	VQ	Online	P	Acc.
Coates et al. (2011)	$1 \cdot 10^3$	✓	×	6	68.6
Ba and Caruana (2014)	$4 \cdot 10^3$	×	✓	-	81.6
Wavelets (Oyallon and Mallat, 2015)	-	×	×	8	82.2
Recht et al. (2019)	$2 \cdot 10^5$	×	×	6	85.6
SimplePatch (Ours)	$1 \cdot 10^4$	✓	✓	6	85.6
SimplePatch (Ours)	$6 \cdot 10^4$	✓	✓	6	86.7
SimplePatch (Ours)	$6 \cdot 10^4$	×	✓	6	**86.9**

Patch K nearest neighbors binary encoding

Ours

Linear classification ImageNet

Method	$\|\mathcal{D}\|$	VQ	P	Depth	Resolution	Top1	Top5
Random (Arandjelovic et al., 2017)	-	$\times$	-	9	224	18.9	-
Wavelets (Zarka et al., 2019)	-	$\times$	32	2	224	26.1	44.7
SimplePatch (Ours)	2.10^3	✓	6	1	64	33.2	54.3
SimplePatch (Ours)	2.10^3	✓	12	1	128	35.9	57.4
SimplePatch (Ours)	2.10^3	$\times$	12	1	128	36.0	**57.6**

When is nearest neighbor meaningful ?

Beyer et al. (1999)

<u>Dimensionality and nearest-neighbors</u>

- « *Under a broad set of conditions, for as few as 10-15 dimensions, the distance to the nearest datapoint approaches the distance to the farthest datapoin*t »

- « *Scenario where high-dimensional nearest neighbors are meaningful occurs when the underlying dimensionality of the data is much lower than the actual dimensionality* »

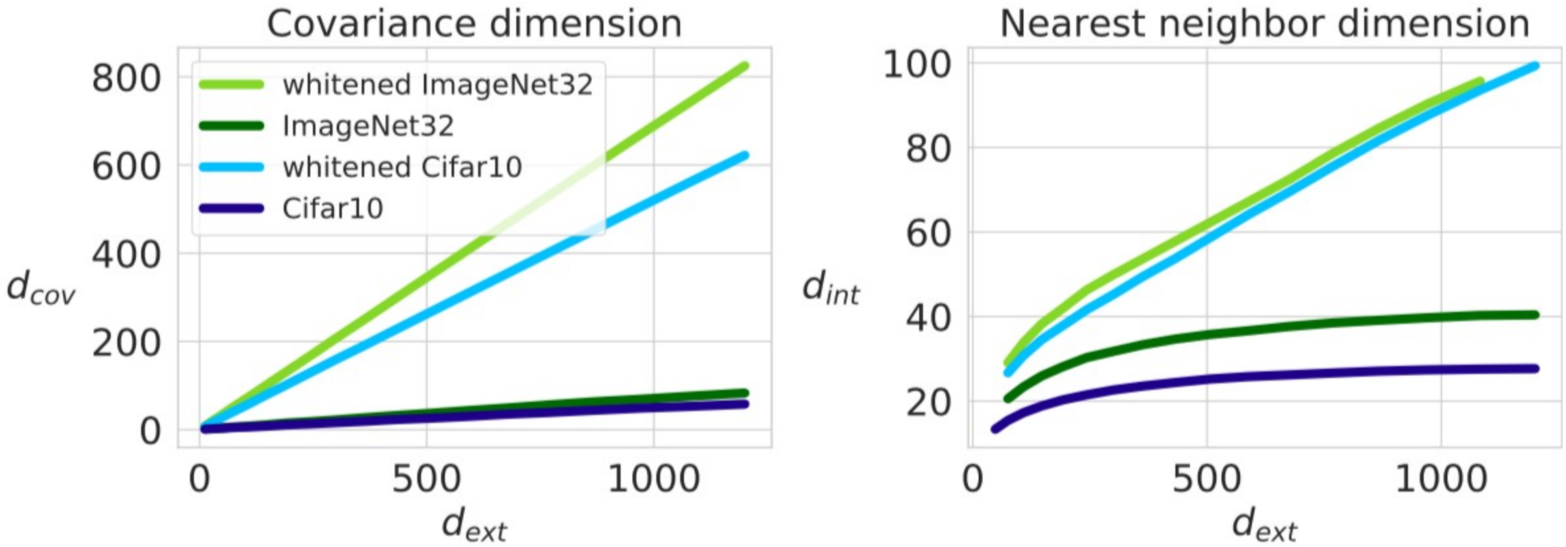
Dimensionality study

Ours

Dimensionality measures

- Covariance dimension : sum of covariance eigenvalues

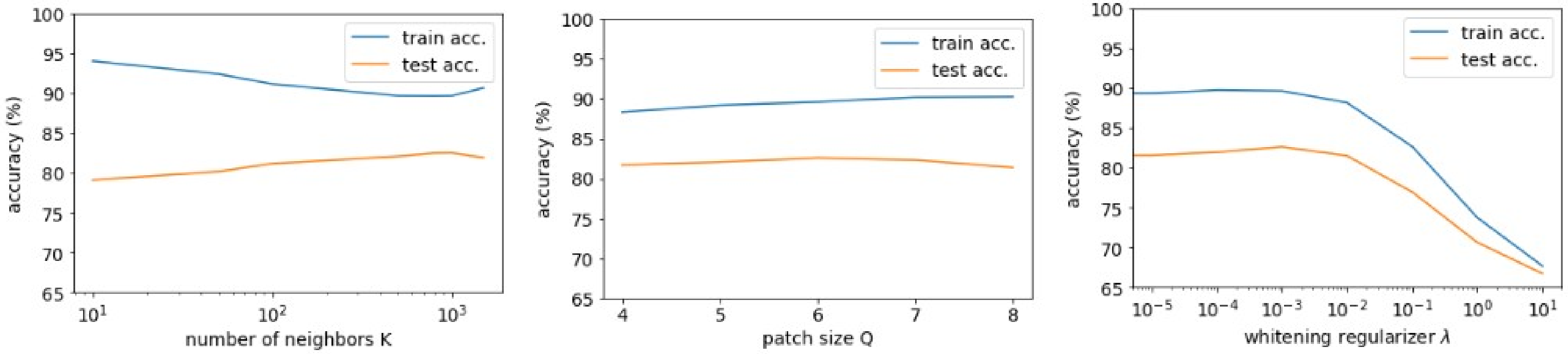
- Nearest-neighbor dimension : $$d_{\text{int}}(p) = \left(\frac{1}{K-1} \sum_{k=1}^{K-1} \log \frac{\tau_K(p)}{\tau_k(p)} \right)^{-1}$$



Patch K nearest neighbors binary encoding

Ours

Ablation study on CIFAR 10



- Large number of neighbors reduces overfitting
- Patch size does not affect the performance
- Whitening $W = (\lambda I + \Sigma)^{-1/2}$ does not need regularization

Patch K nearest neighbors binary encoding

Ours

Remarks

- Competitive performance with shallow classifiers
- Form of low-dimensionality in natural image patches
- Whitening is key aspect
- Relatively stable with large number of neighbors, and the derivative of the output w.r.t input is zero

Neural style transfer with artists

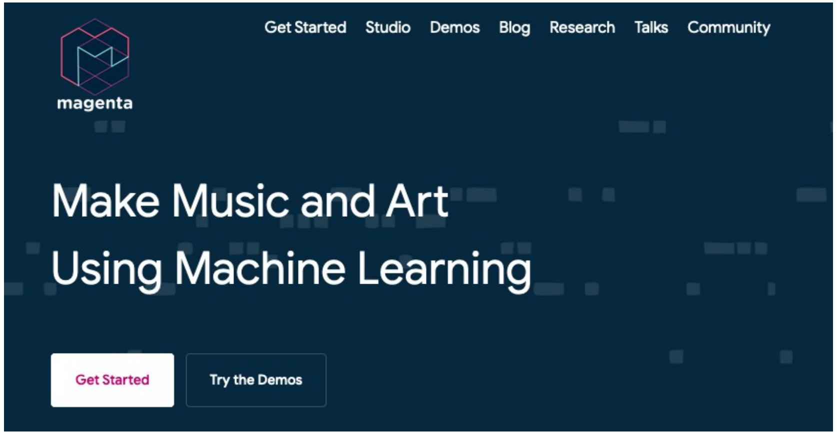
T. Kerdreux, L. Thiry, E. Kerdreux

https://arxiv.org/abs/2003.06659

Machine learning for creativity

<u>Software for the creative industry</u>

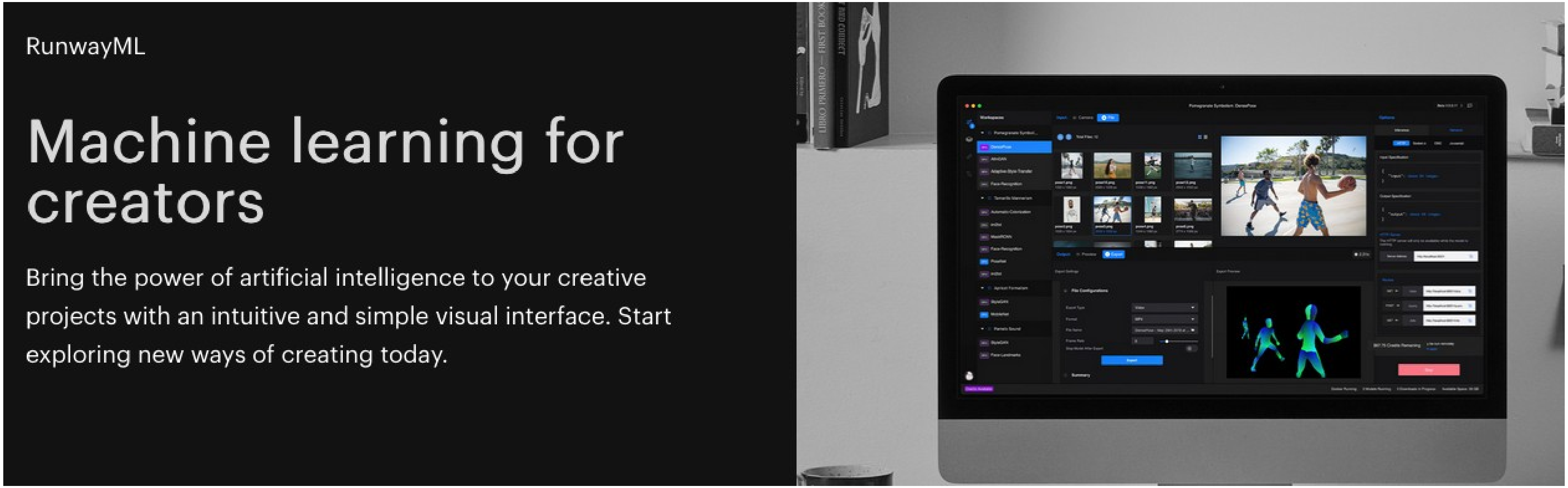
- Photoshop (Adobe)
- Magenta Tensorflow (Google)



Machine learning for creativity

Software for the creative industry

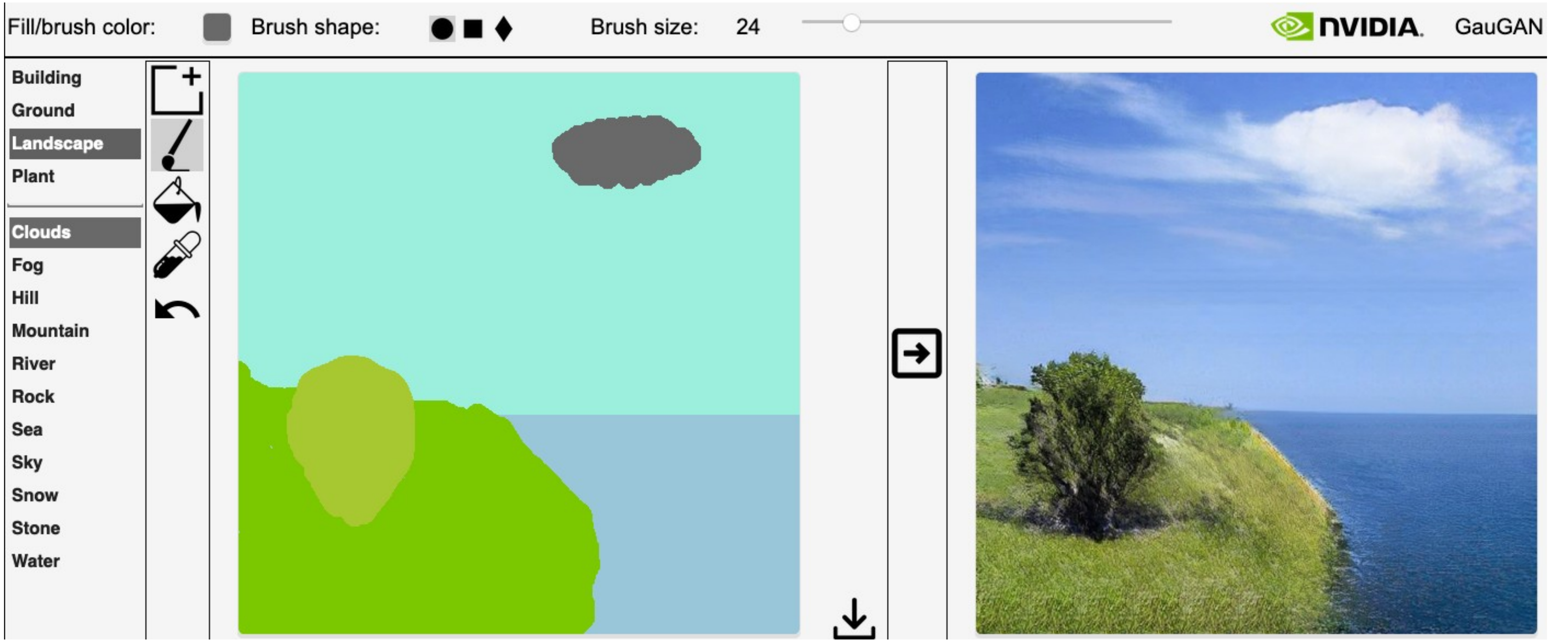
- Runaway ML



Machine learning for creativity

Software for the creative industry

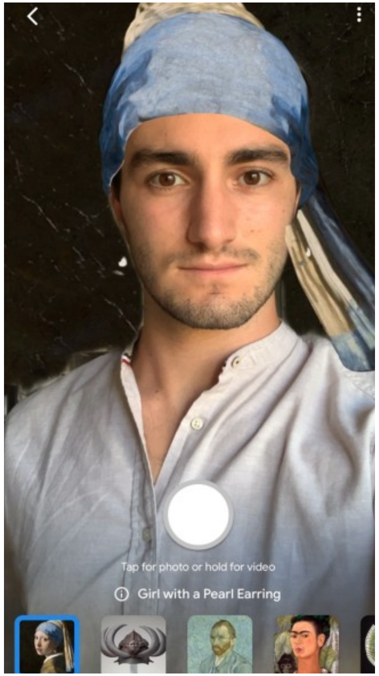
- NVIDIA GauGAN

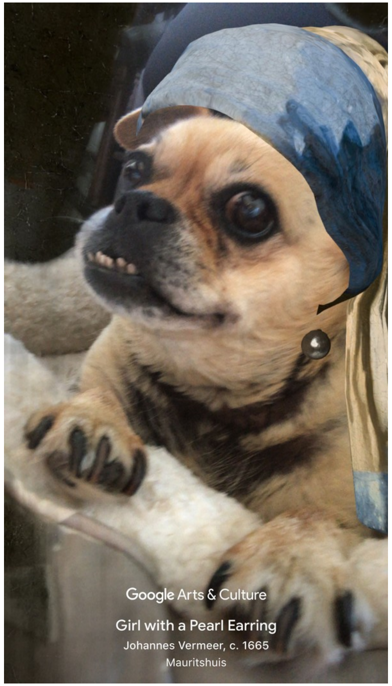


Machine learning for art « gamification »

<u>Interactive software</u>

- Google art and culture app





« Art » with machine learning

Artwork generation with GAN

The Butcher's son, Lumen Prize 2019

« Generated with pornography images »

MARIO KLINGEMANN

CHF7,750.00

The Butcher's Son, 2017
Hahnemühle Paper Museum Etching 350 gms
Work size: 76 x 50 cm
Unique edition + 1 AP
© Mario Klingemann courtesy Onkaos
Framed

« Art » with machine learning

Dear Glenn. Yamaha Corporation , 2019

This is a project dubbed Dear Glenn, a project inspired by his unique creative style and launched to explore the future of music through the use of artificial intelligence.
Glenn Gould was known for his devotion to recording with digital media and an interest in rethinking the relationship between performer and audience.
The project to develop this system has been dubbed “Dear Glenn” as a tribute to the artist’s attitude, which was the inspiration for the idea behind the project.

« Art » with machine learning

The next Rembrandt.

Microsoft, ING, TU Delft, 2016.

« To distill the artistic DNA of Rembrandt, an extensive database of his paintings was built and analyzed, pixel by pixel. »

Artification

Shapiro (2004), Shapiro et Heinich (2012)

« Art is not a given and cannot be defined once and for all. It is a construct and the result of social processes that are located in time and place. »

« Art emerges over time as the sum total of institutional activities, everyday interactions, technical implementations, and attributions of meaning. »

Style Transfer

- Transfering the semantic content of an image and the style of another image into a new one

Neural Style Transfer

Gatys et Al. (2015)

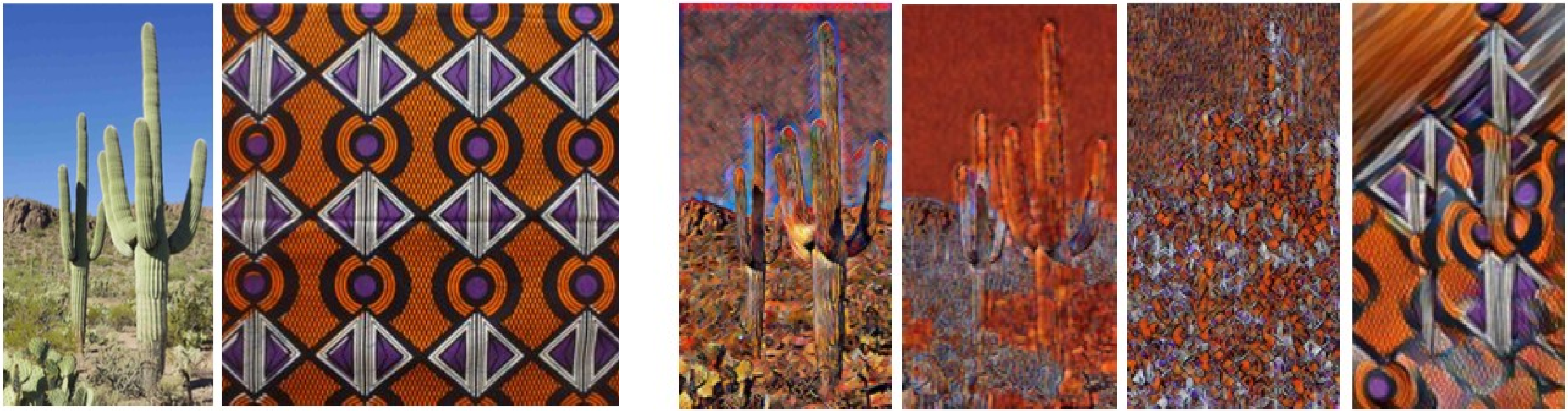
- Pretrained convolutional network Φ
- Content image I_c and style image I_s
- Gram matrix of the style features $G[\Phi(I_s)]$
- Optimization problem :

$$\min_I \|\Phi(I) - \Phi(I_c)\|^2 + \alpha \|G[\Phi(I)] - G[\Phi(I_s)]\|^2$$

Neural Style Transfer

Gatys et Al. (2015)

Other style transfer methods



(a) content (b) style (c) AdaIn (d) MST (e) WCT (f) STROTSS

How do they compare with each other ?

Photo-painting pairs for qualitative evaluation

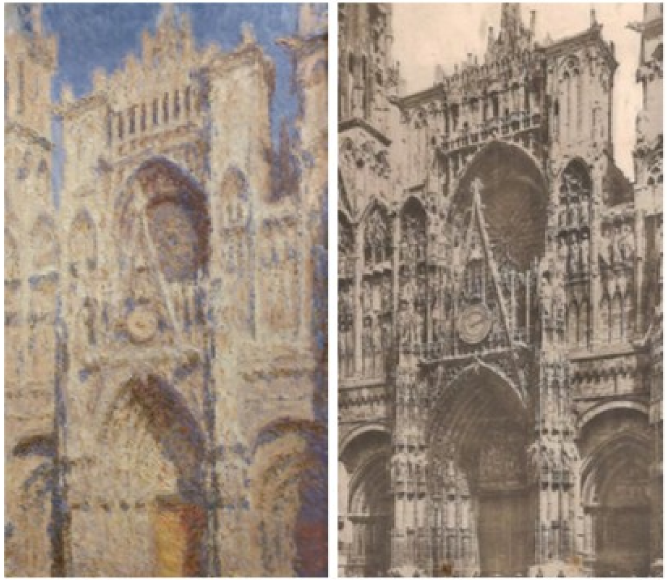
Ours

- Claude Monet Rouen Cathedral series

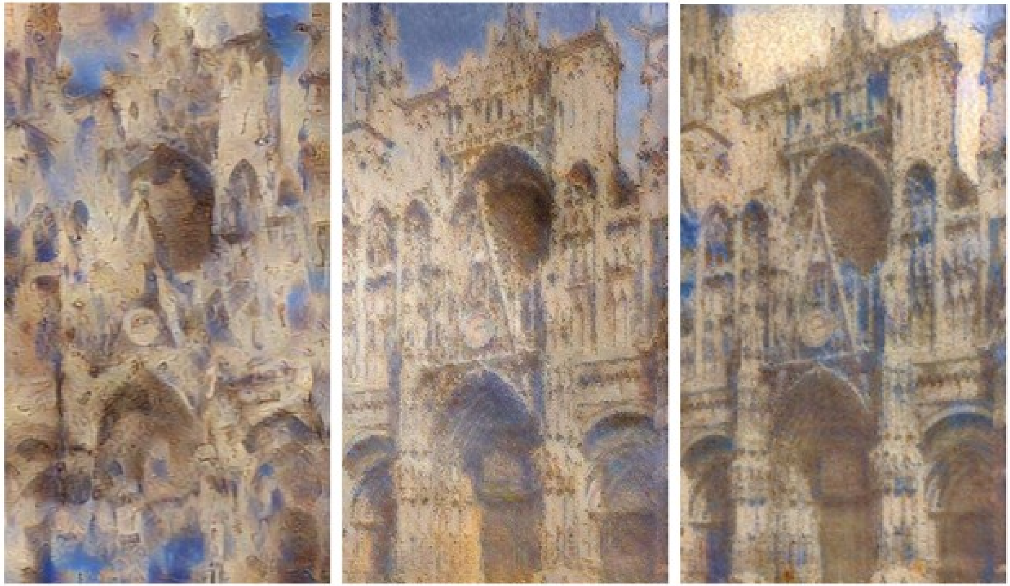
Photo-painting pairs for qualitative evaluation

Ours

- Photo-painting alignment for style transfer



(a) Monet painting (b) photograph

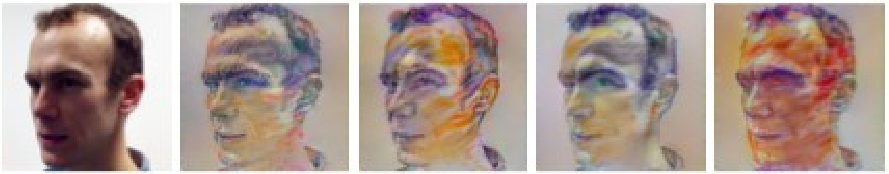


(d) WCT (e) STROTSS (f) Gatys

Interactive experiements

Ours

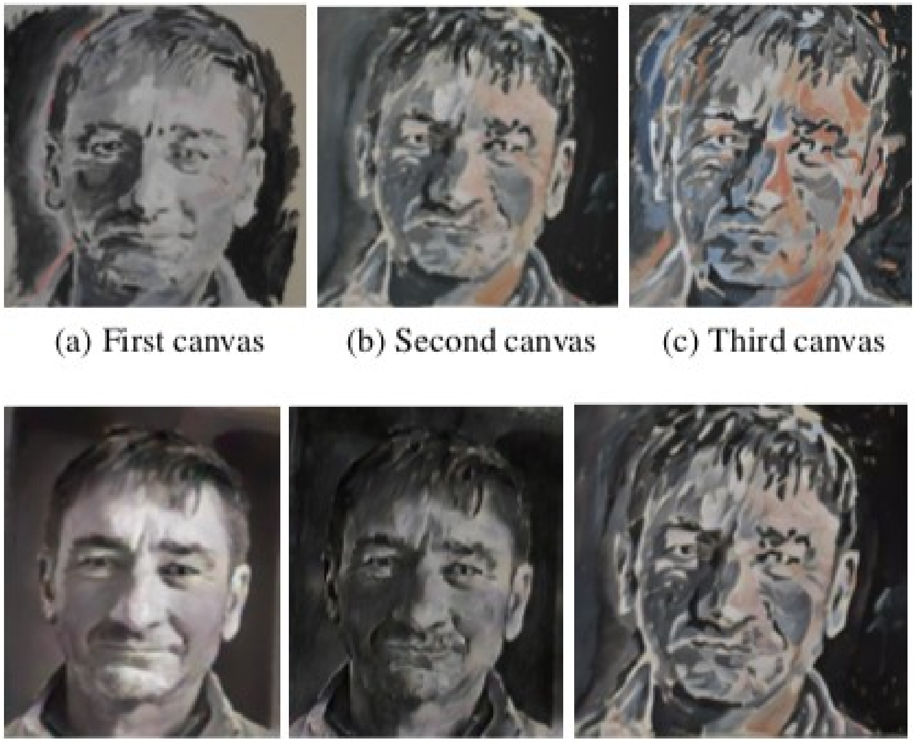
(left) final canvas (right) steps 1, 3, 5, 7



Original photographic and projections after steps 1, 3, 5, 7.

Interactive experiments

Ours



(a) First canvas (b) Second canvas (c) Third canvas

(d) Projection 1 (e) Projection 2 (f) Projection 3

Interactive experiments

Ours

Machine learning and artistic field

- *Artification* and societal consequences.

- Artwork offer evaluation of what is acheived by the current methods
- Artist interaction with the an algorithm : new source of inspiration rather than machine creativity
- Different perspective on research