Efficient representations with Auto-Encoders and Max-Approximations to morphological operators

Supervisors: Jesus ANGULO, Santiago VELASCO-FORERO, Samy BLUSSEAU, Isabelle BLOCH

école normale supérieure paris–saclay

Bastien PONCHON
December 12th 2018

MINES ParisTech

Agenda

01 - Introduction

Representation Learning and Part-Based representation

Representation Learning:

- Learning an underlying structure/process explaining the M input images $\mathbf{x}^{(i)} \in E^N, i \in [1, M]$ (of N pixels), that can somehow be represented as a set of latent features $\mathbf{h}^{(i)} \in E^k, i \in [1, M]$ in a space of dimension k
- If the data points live on a manifold of lesser dimension than the original space: $k < N$

Sparse coding and dictionary learning:

- The input images is assumed to be well represented as a weighted linear combination of a few elements from a dictionary, called the **atom images,** $\mathbf{w}_j \in E^N, j \in [1, k]$:

$$\forall i \in [1, M], \mathbf{x}^{(i)} \approx \sum_{j=1}^{k} h_{i,j}\mathbf{w}_j = \mathbf{h}^{(i)}\mathbf{W} = \hat{\mathbf{x}}^{(i)}$$

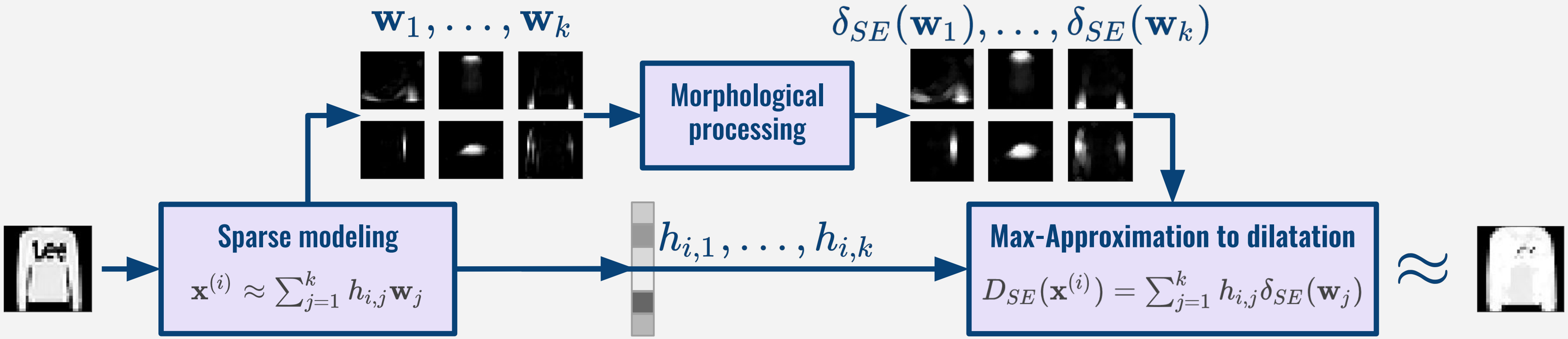
Part-based representation:

- Introduced by Lee and Seung in their 1999 work about NMF: atom images representing localized features corresponding with intuitive notions of the parts of the input image family.

Max-Approximation to Morphological Operators

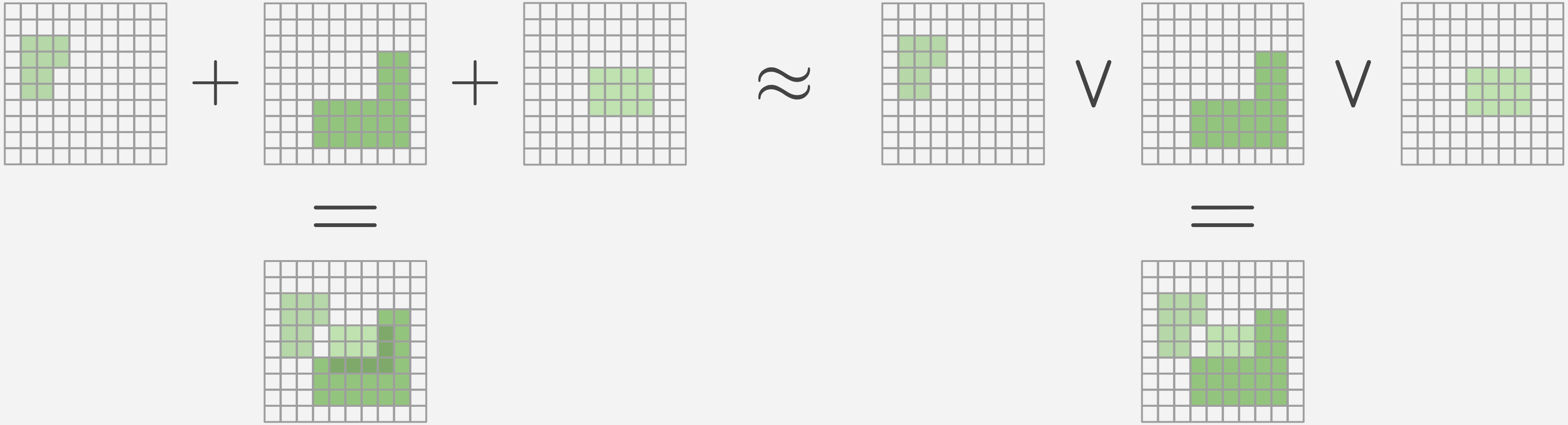
"*Sparse mathematical morphology using non-negative matrix factorization*" , **Angulo, Velasco-Forero 2017:** Exploring how image sparse representations can be useful to efficiently calculate approximations to morphological operators, applied to a whole set of images.

Sparse Max-Approximation to gray-level dilation:



Sparse Max-Approximation to erosion, opening, closing...

Motivation for Non-Negative and Sparse representation



$$\forall i \in [1, M], \qquad \forall (j,l) \in [1,k]^2, h_{i,j}\mathbf{w}_j \bigwedge h_{i,l}\mathbf{w}_l \approx 0$$

$$\implies \bigvee_{j \in [1,k]} h_{i,j}\mathbf{w}_j \approx \sum_{j=1}^{k} h_{i,j}\mathbf{w}_j$$

$$\implies D_{SE}(\mathbf{x}^{(i)}) \approx \delta_{SE}(\mathbf{x}^{(i)})$$

Objectives and Motivations

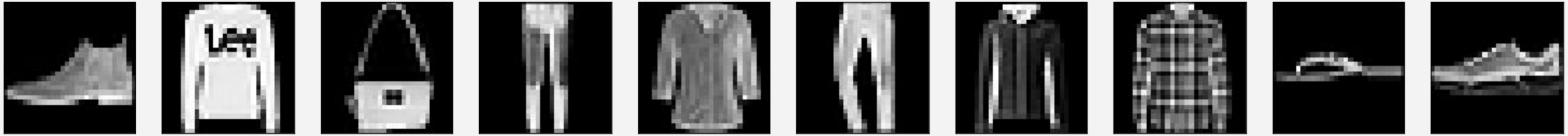
Using Neural Networks to learn a non-negative and sparse part-based representation:

- No need to re-train the model to encode new, previously unseen, images, unlike NMF.
- Ability to approximate the application of various morphological operators (dilations, erosions, openings, closings, morphological gradient, black top-hat, etc.) to an unlimited number of images by applying these operators only to the k atom images.

The most intuitive and common way to perform representation learning in the Deep Learning paradigm is to use Auto-Encoders.

Evaluation and Data of the Proposed Models

The Fashion-MNIST database of images:



Evaluation criteria of the learned representation of a test set of images not used to train the model (except for the NMF):

- **Approximation error of the representation**: mean-squared error between the original input images and their approximation by the learned representation
- **Max-approximation error to the dilation** by a disk of radius 1: mean-squared error between the max-approximation to the dilation and the dilation of the original input images
- **Sparsity of the encoding**, measured using the metric introduced by Hoyer (2004)
- **Classification Accuracy** of a linear Support-Vector Machine, taking as input the encoding of the images.

02 - Non-Negative Matrix Factorization

General Presentation

"*Learning the parts of objects by non-negative matrix factorization*", **Lee and Seung, 1999:**

- Matrix factorization algorithm:

  $$\hat{\mathbf{X}} = \mathbf{HW} \approx \mathbf{X}$$

  with $\mathbf{X} \in E^{M \times N}$ the **data matrix** containing the M images of N pixels, as row vectors

  $\mathbf{W} \in E^{k \times N}$ the **dictionary matrix**, containing the k atom images as row vectors

  $\mathbf{H} \in E^{M \times k}$ the **encoding matrix**, containing the representation of each of the images as row vectors

- Proven to actually recover the parts of the images if the data set of images is a **separable factorial articulation family:**
  - Each image actually generated by a linear combination of positive atom images associated with non-negative weight
  - All atom images have separated supports.
  - All different combinations of parts are exhaustively sampled in the data set of images.

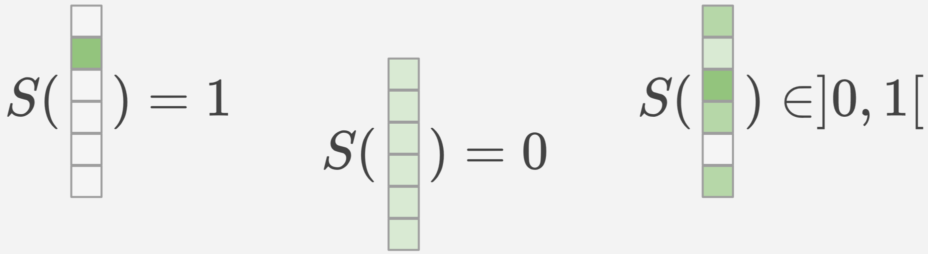
Addition of sparsity constraints (Hoyer 2004)

“*Non-negative matrix factorization with sparseness constraints*”, Hoyer, 2004:

- Enforcing sparsity of the encoding and/or of the atoms of the NMF representation: most coefficients taking values close to zero, while only a few take significantly non-zero values.

Sparsity measure of vector $\mathbf{v} \in E^d$:

$$S(\mathbf{v}) = \frac{\sqrt{d} - \frac{||\mathbf{v}||_1}{||\mathbf{v}||_2}}{\sqrt{d} - 1} \in [0, 1]$$

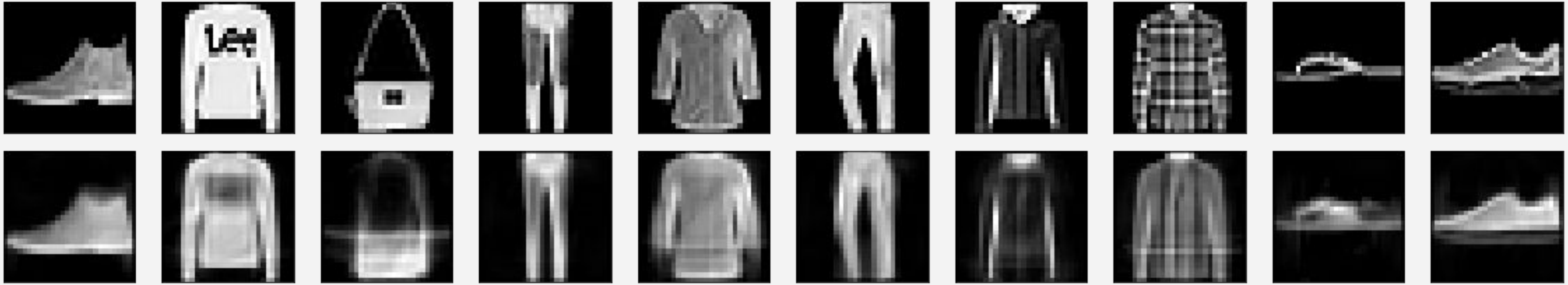


After each update of $\mathbf{H}$ and $\mathbf{W}$ in the NMF algorithm, the encodings and atoms are projected on the space verifying:

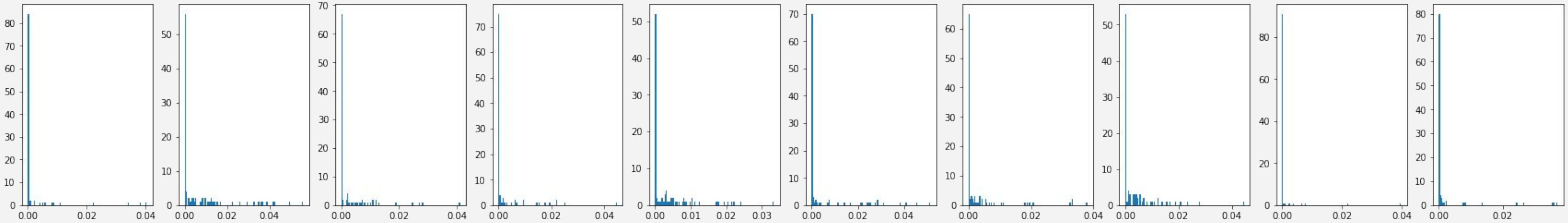
$$S(\mathbf{h}^{(i)}) = S_h, \forall i \in [1, M]$$
$$S(\mathbf{w}_j) = S_w, \forall j \in [1, k]$$

Results - $S_h = 0.6$

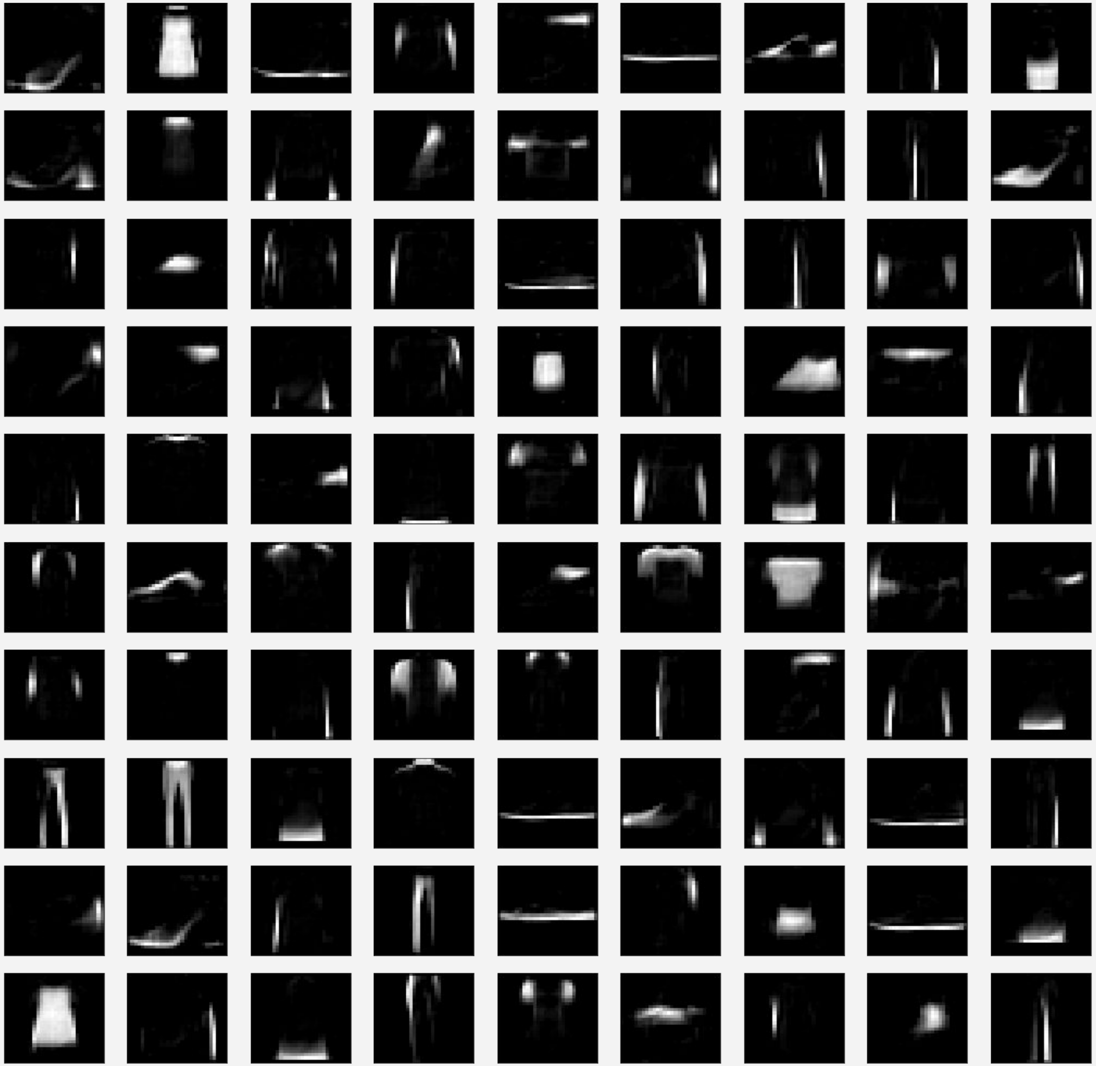
Original images and reconstruction - *Reconstruction error: 0.0109*



Histogram of the encodings - *Sparsity metric: 0.650*

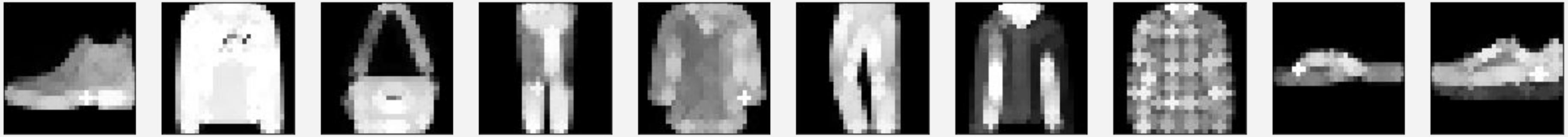


Atom images of the representation

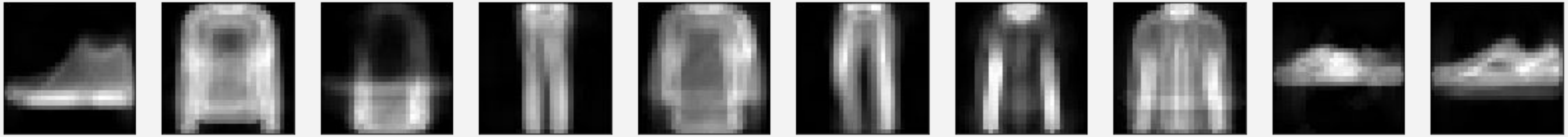


Results - Max-Approximation to dilation

Dilation of the original images by a disk of radius 1



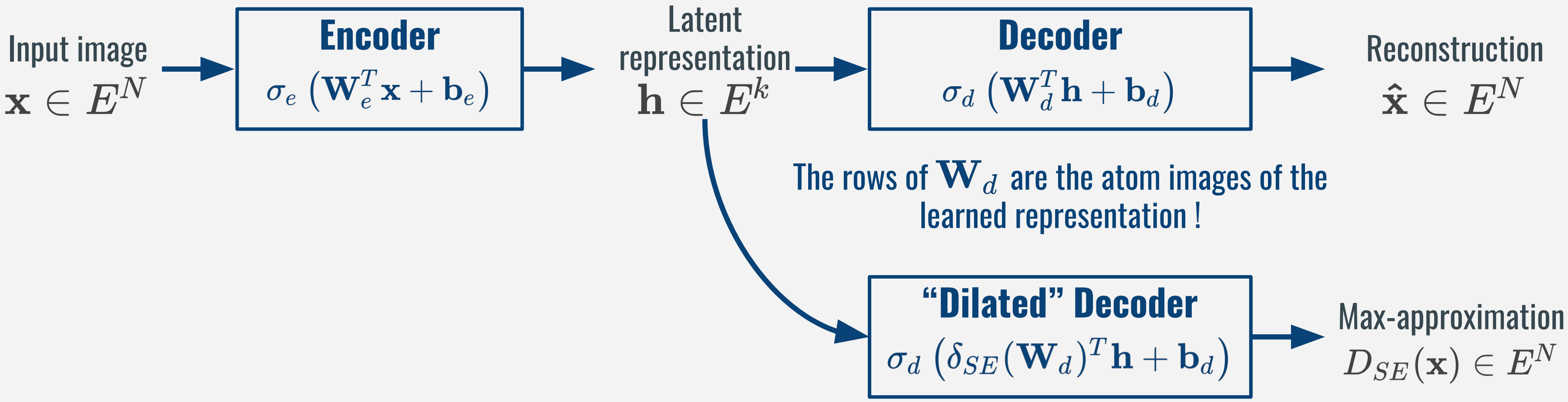
Max-approximation to the dilation by a disk of radius 1 - *Max-approximation error: 0.107*



	Dilation	Max-Approximation
Computation Time on 10000 images *(in s)*	1.039	0.075

03 - Part-based Representation using Asymmetric Auto-Encoders

Shallow Auto-Encoders



Auto-encoder loss function, minimized during training:

$$L_{AE} = \frac{1}{M}\sum_{i=1}^{M} L(\mathbf{x}^{(i)}, \hat{\mathbf{x}}^{(i)})$$ where $L(\mathbf{x}^{(i)}, \hat{\mathbf{x}}^{(i)})$ is the reconstruction error (MSE)

An Asymmetric Auto-Encoder

Input image $\mathbf{x} \in E^N$ → **infoGAN** → Latent representation $\mathbf{h} \in E^k$

→ **Decoder** $\sigma_d \left(\mathbf{W}_d^T \mathbf{h} + \mathbf{b}_d \right)$ → Reconstruction $\hat{\mathbf{x}} \in E^N$

→ **"Dilated" Decoder** $\sigma_d \left(\delta_{SE} (\mathbf{W}_d)^T \mathbf{h} + \mathbf{b}_d \right)$ → Max-approximation $D_{SE}(\mathbf{x}) \in E^N$

"InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets", Chen *et al.* 2016

- Two 2D convolutional layers
- Two fully connected layers

Motivations:

- Designed for a representation learning task on MNIST data set.
- Simple architecture.
- Use of convolutional layers, well adapted to computer vision tasks.
- Use of widely adopted state of the art techniques in deep learning: batch-normalization, leakyRELU, etc.

Enforcing the Sparsity of the Encoding

Regularization of the auto-encoder:

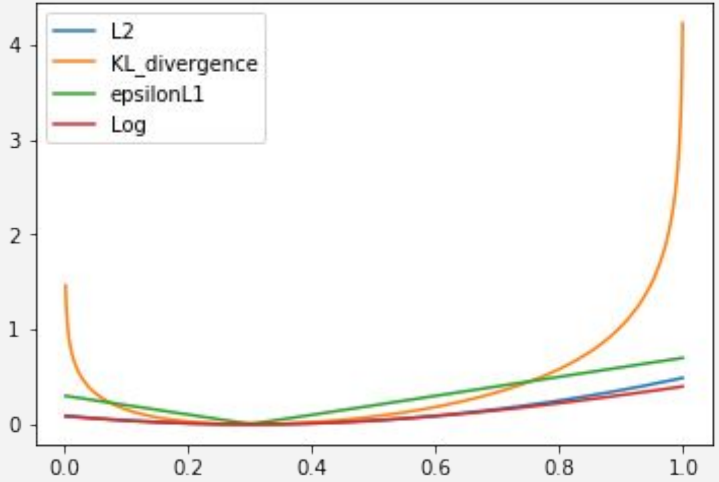
Sparsity constraint

$$L_{AE} = \frac{1}{M}\sum_{i=1}^{M} L(\mathbf{x}^{(i)}, \hat{\mathbf{x}}^{(i)}) + \beta \sum_{j=1}^{k} S(\frac{1}{M}\sum_{i=1}^{M} h_j^{(i)}, p)$$

Penalizes a deviation of the **expected activation of each hidden unit** from a (low) **fixed level**

Various choices for the sparsity-regularization function:

$$S_{KL}(t_j, p) = p \log \frac{p}{t_j} + (1-p)\log\frac{1-p}{1-t_j}$$

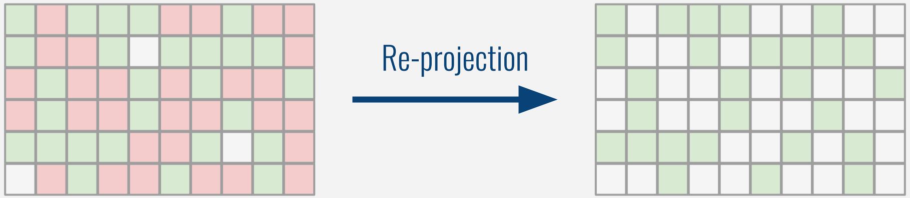


Enforcing Non-Negativity of the Atoms of the Dictionary

Re-Projection on the positive orthant:

- Non-Parametric constraint
- Ensured non-negativity

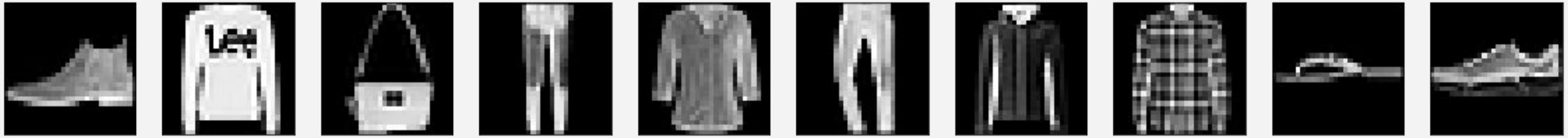
After each iteration of the optimization algorithm (e.g.: Stochastic Gradient Descent):



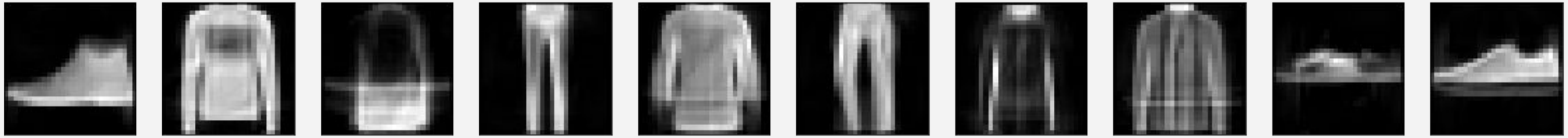
04 - Results and Conclusion

Reconstructions

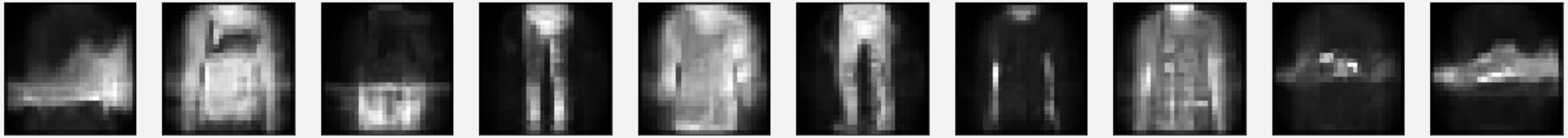
Original Images



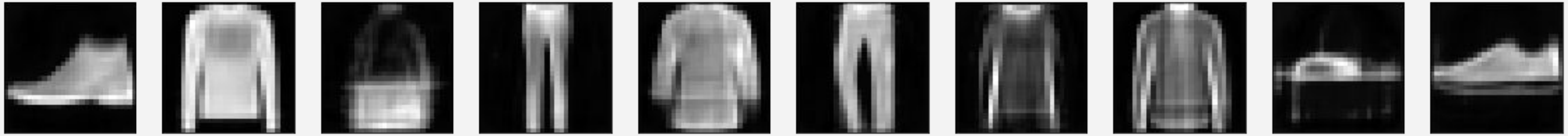
Sparse-NMF (Hoyer 2004) *(S_h=0.6)* - *Reconstruction error: 0.0109*



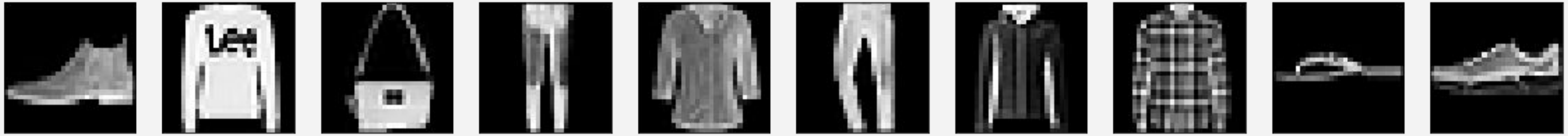
NNSAE (Lemme et al. 2011) - *Reconstruction error: 0.0514*



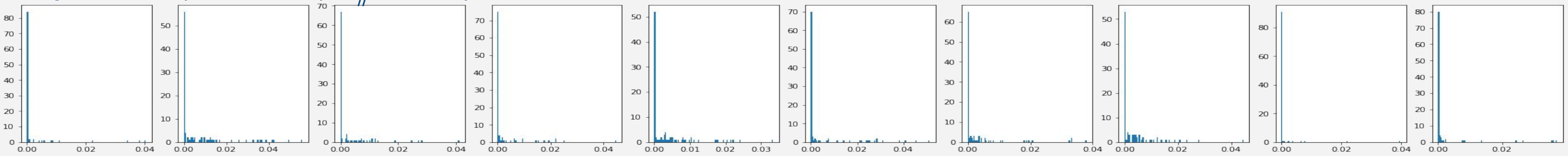
Sparse, Non-Negative Asymmetric AE *(p=0.05, beta=0.005)* - *Reconstruction error: 0.0125*



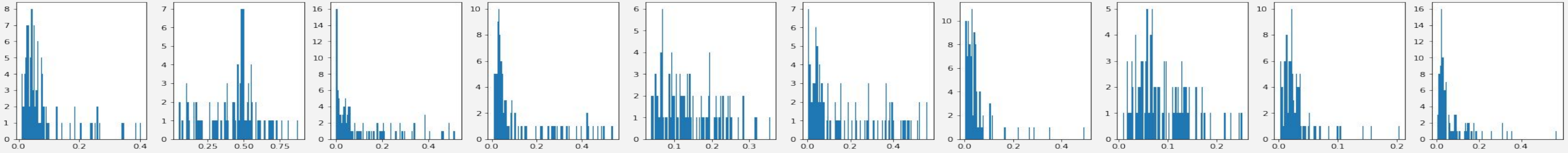
Encodings



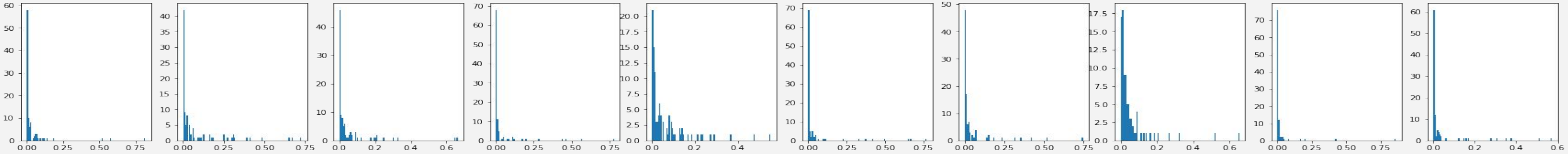
Sparse-NMF (Hoyer 2004) *(S_h=0.6)* - *Sparsity: 0.650*



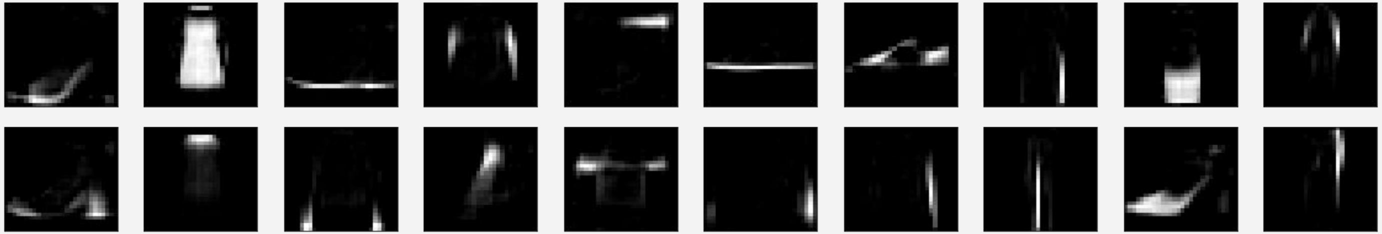
NNSAE (Lemme et al. 2011) - *Sparsity: 0.220*



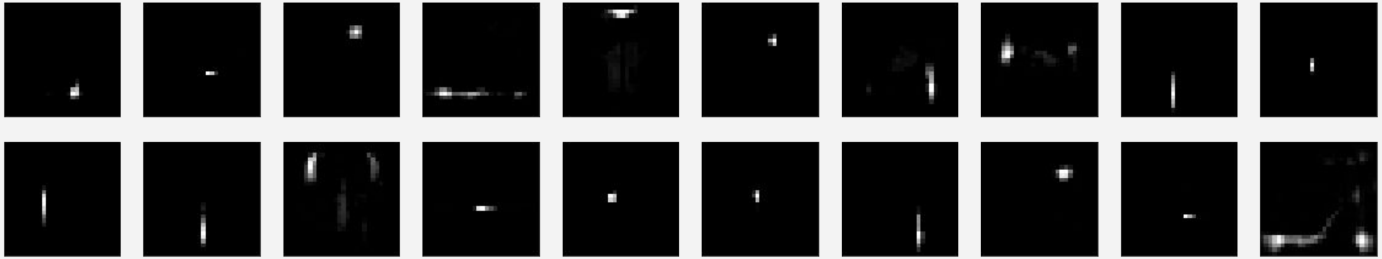
Sparse, Non-Negative Asymmetric AE *(p=0.05, beta=0.005)* - *Sparsity 0.615*



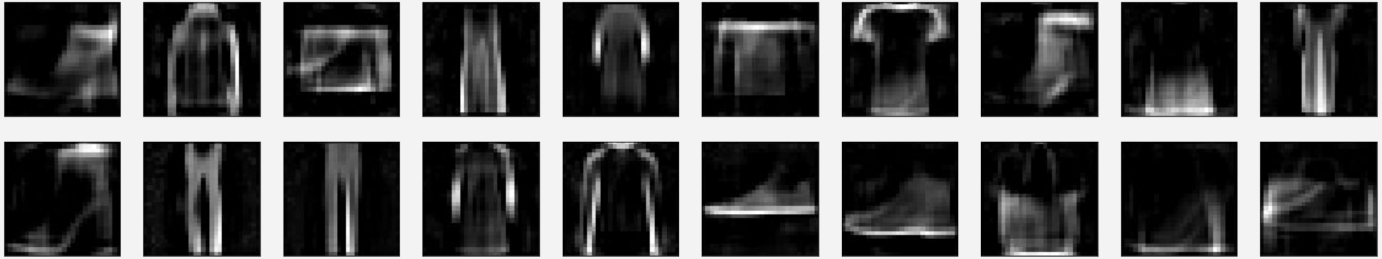
Atoms



Sparse-NMF (Hoyer 2004) *(S_h=0.6)* -
Hoyer Sparsity : 0.762



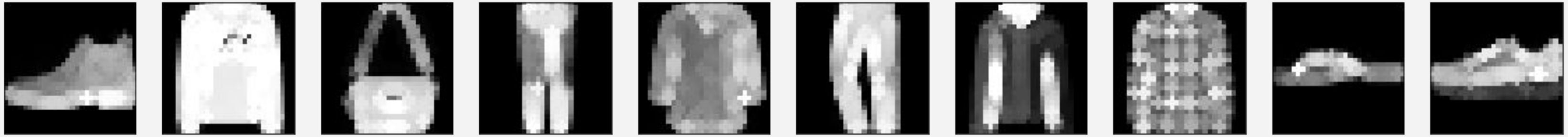
NNSAE (Lemme et al. 2011) -
Sparsity: 0.892



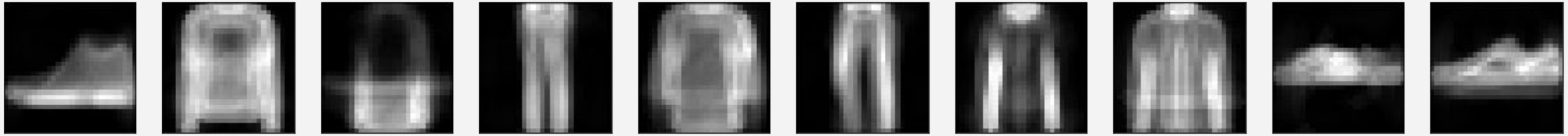
Sparse, Non-Negative Asymmetric AE *(p=0.05, beta=0.005)* -
Sparsity: 0.4703

Max-Approximations to dilation

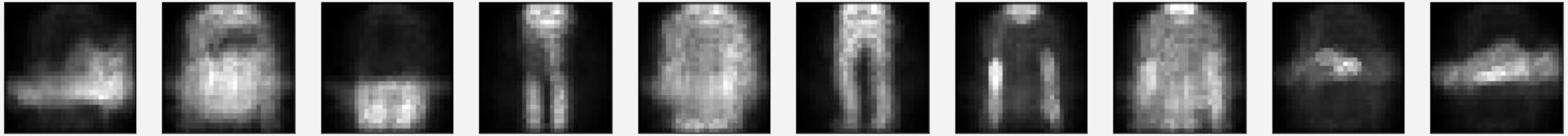
Dilation of Original Images



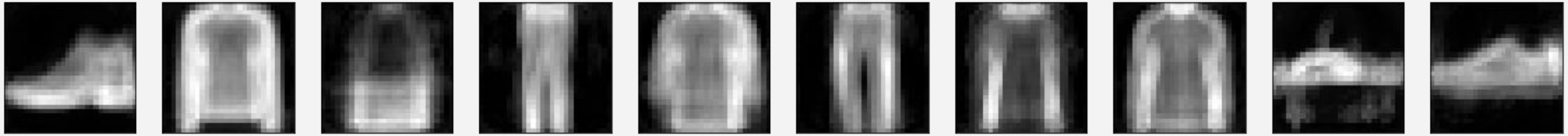
Sparse-NMF (Hoyer 2004) *(S_h=0.6) - Max-Approximation error: 0.107*



NNSAE (Lemme et al. 2011) - *Max-Approximation error: 1.123*



Sparse, Non-Negative Asymmetric AE *(p=0.05, beta=0.005)* - *Max-Approximation error: 0.123*



Perspectives and other works

- Morphological pre-processing using Additive Morphological Decomposition
- Replacing the linear decoder with a Max-Plus Dense Layer