Advanced School in Artificial Intelligence

Autoencoders

Ing. Zese Riccardo

riccardo.zese@unife.it

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regione Emilia-Romagna

FERRARIAE UNIVERSITAS · EX LABORE FRUCTUS · 13 91

Università degli Studi di Ferrara

Outline

- Introduction to Python
- Introduction to Neural Networks
- Convolutional NN
- Recurrent NN
- Autoencoders and self supervised learning

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Outline

- Introduction to Python
- Introduction to Neural Networks
- Convolutional NN
- Recurrent NN
- Autoencoders and self supervised learning

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Supervised vs Unsupervised Learning

Supervised Learning

- **Data:** (x, y)
  x is data, y is label
- **Goal:** Learn a function to map x -> y
- **Examples:** Classification, regression, object detection, semantic segmentation, image captioning, etc.

Unsupervised Learning

- **Data:** x
  Just data, no labels!
- **Goal:** Learn some underlying hidden structure of the data
- **Examples:** Clustering, dimensionality reduction, feature learning, density estimation, etc.

Supervised vs Unsupervised Learning

Collect data is cheaper

Supervised Learning

- **Data:** (x, y)
  x is data, y is label
- **Goal:** Learn a function to map x -> y
- **Examples:** Classification, regression, object detection, semantic segmentation, image captioning, etc.

Unsupervised Learning

- **Data:** x
  Just data, no labels!
- **Goal:** Learn some underlying hidden structure of the data
- **Examples:** Clustering, dimensionality reduction, feature learning, density estimation, etc.

Regione Emilia-Romagna

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

Supervised vs Unsupervised Learning

Supervised Learning

- **Data:** (x, y)
  x is data, y is label
- **Goal:** Learn a function to map x -> y
- **Examples:** Classification, regression, object detection, semantic segmentation, image captioning, etc.

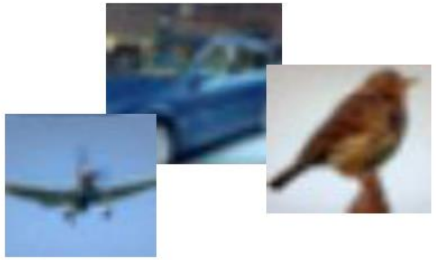
Unsupervised Learning

- **Data:** x
  Just data, no labels!
- **Goal:** Learn some underlying hidden structure of the data
- **Examples:** Clustering, dimensionality reduction, feature learning, density estimation, etc.

Generative Models

Generative Models

- Given training data, generate new samples from same distribution



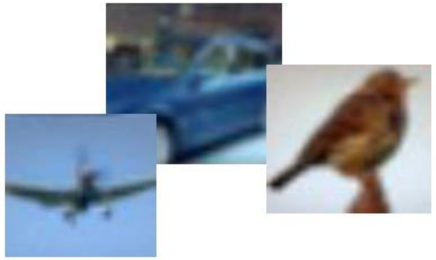
Training data $\sim p_{data}(x)$

Generated data $\sim p_{model}(x)$

Want to learn $p_{model}(x)$ as much close as possible to $p_{data}(x)$

Generative Models

- Given training data, generate new samples from same distribution



Training data $\sim p_{data}(x)$

Generated data $\sim p_{model}(x)$

Want to learn $p_{model}(x)$ as much close as possible to $p_{data}(x)$

Addresses *density estimation*, a core problem in unsupervised learning

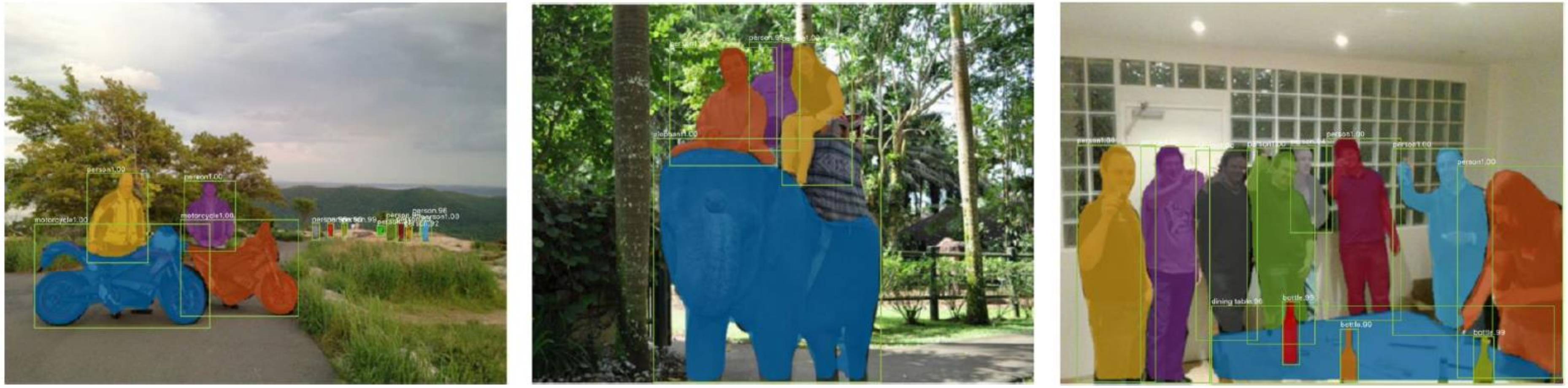
Several flavours:

- **Explicit density estimation**: explicitly define and solve for $p_{model}(x)$
- **Implicit density estimation**: learn model that can sample from $p_{model}(x)$ w/o explicitly defining it

Motivations

- Generate realistic samples
- Generative models of time-series data can be used for simulation and planning
- Training generative models can also enable inference of latent representations that can be useful as general features

Motivations



He et al, "Mask R-CNN", arXiv 2017
Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Motivations

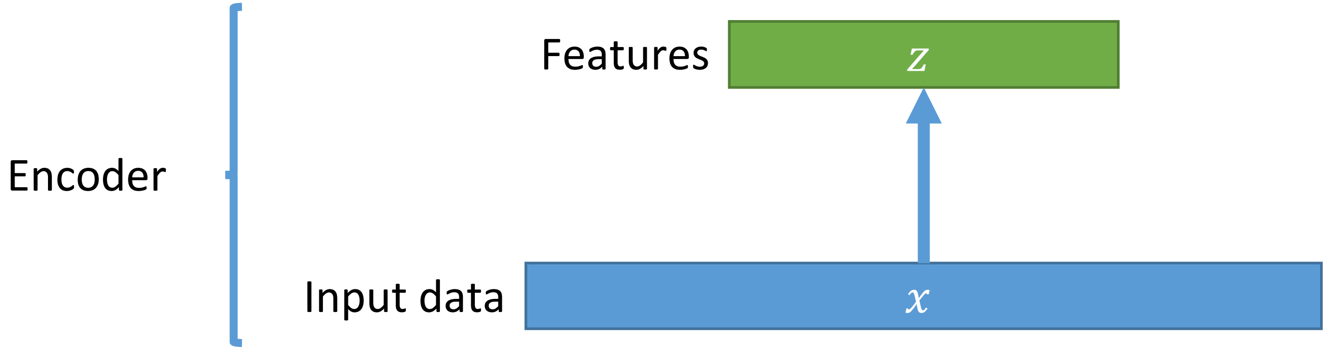
… and many more!

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

He et al, "Mask R-CNN", arXiv 2017
Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.

Autoencoders

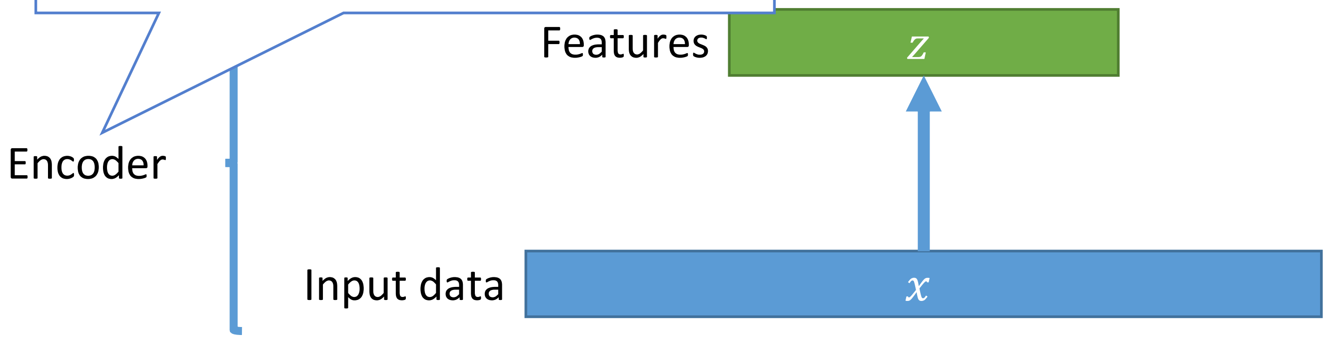
- Unsupervised approach for learning a lower-dimensional feature representation from unlabeled training data



Autoencoders

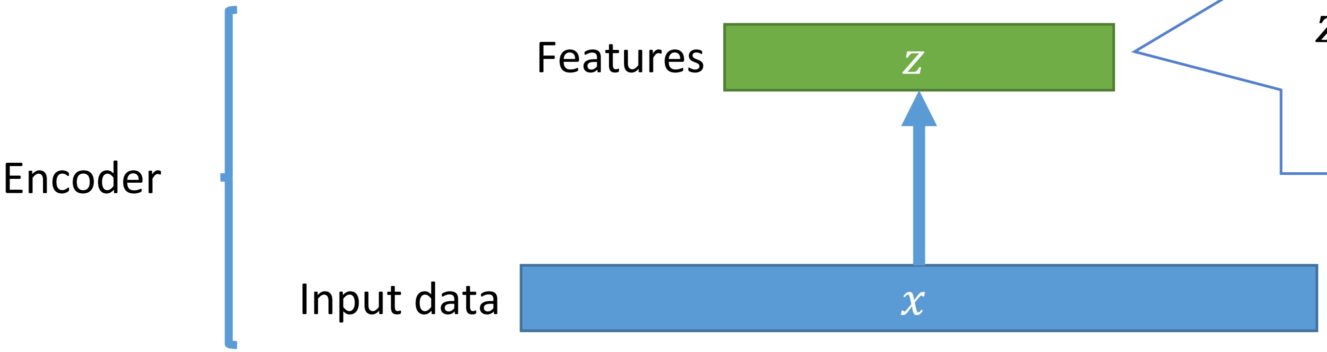
- Unsupervised approach for learning a lower-dimensional feature representation from unlabeled training data

Originally: Linear + nonlinearity (sigmoid)
Later: Deep, fully-connected
Then: ReLU CNN



Autoencoders

- Unsupervised approach for learning a lower-dimensional feature representation from unlabeled training data



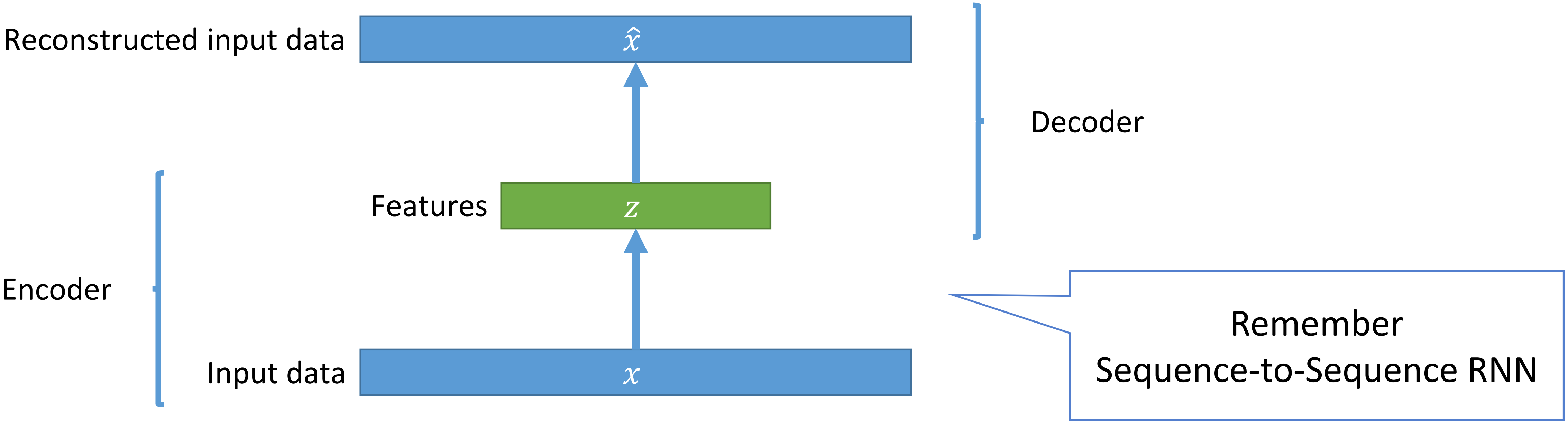
z usually smaller than x:
dimensionality reduction
in order to have features that capture meaningful factors of variation in data.
z represents the most important features in x

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

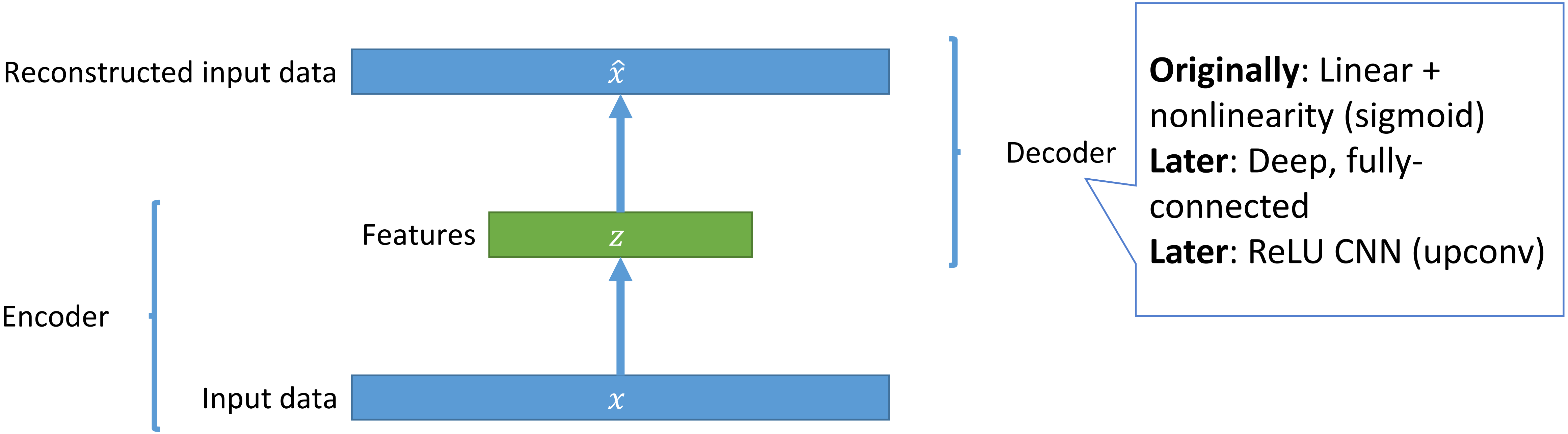
Autoencoders

- This feature representation is trained so that features can be used to reconstruct original data → **Autoencoding** - encoding itself

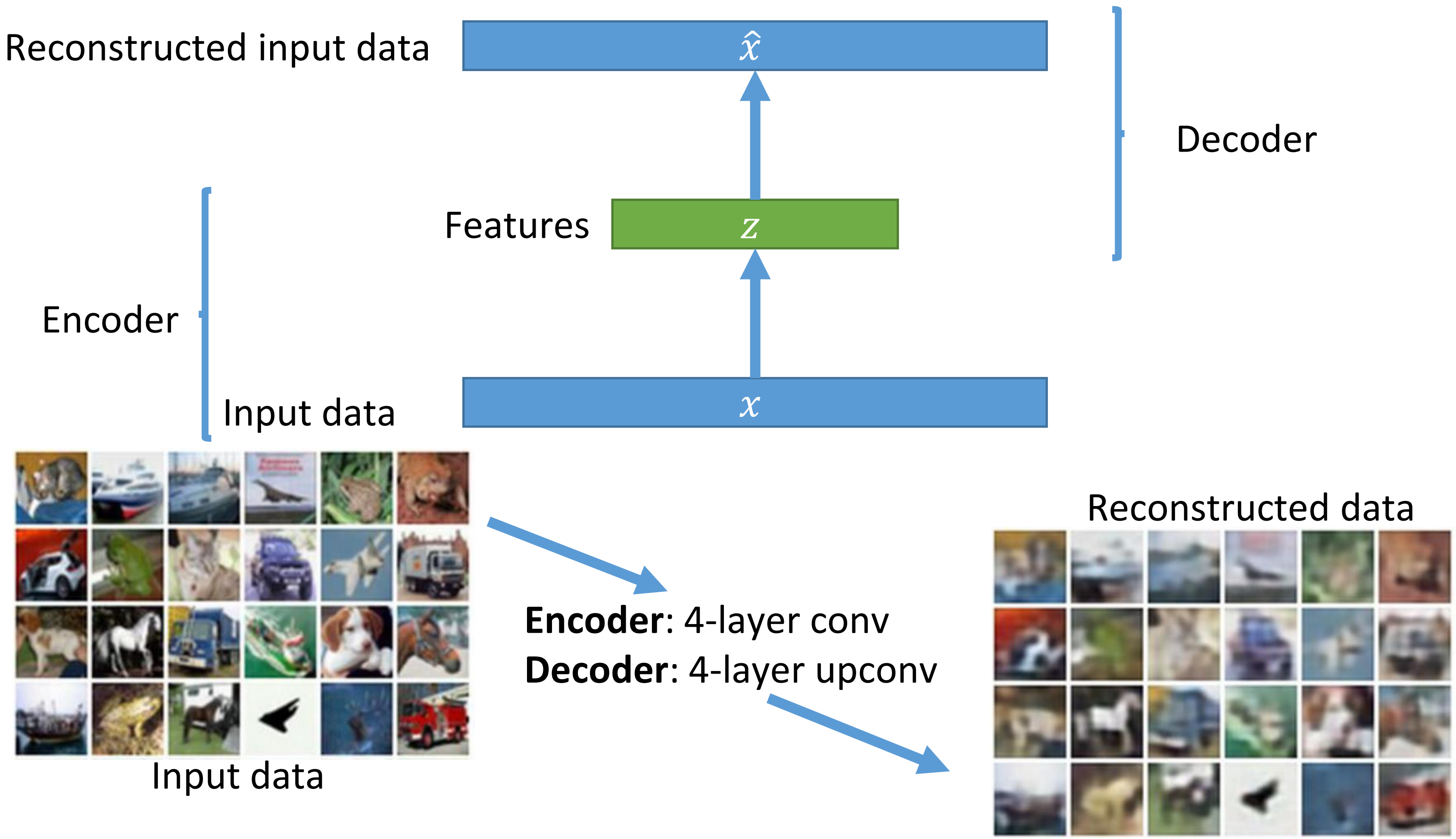


Autoencoders

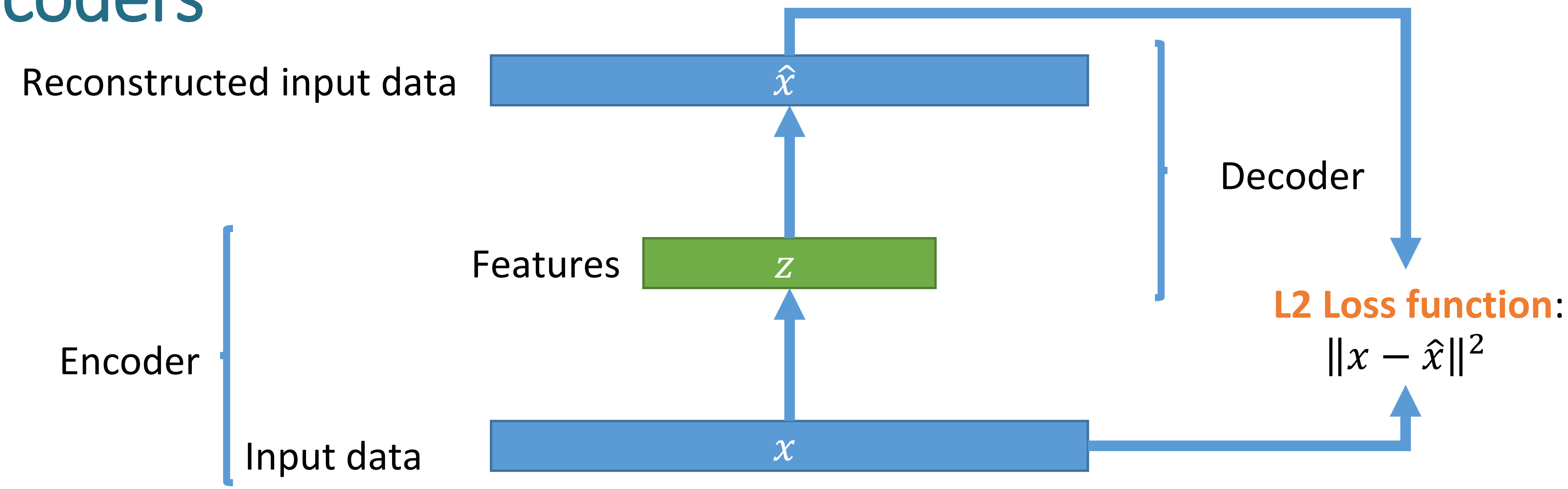
- This feature representation is trained so that features can be used to reconstruct original data
  → **Autoencoding** - encoding itself

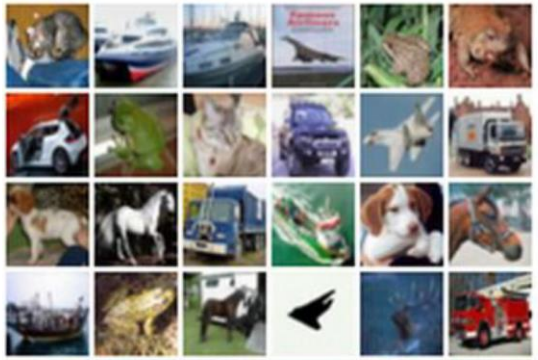


Autoencoders



Autoencoders

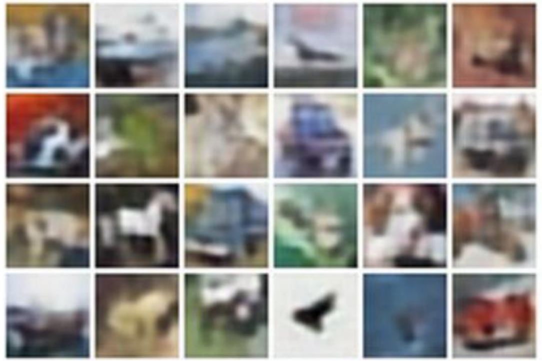




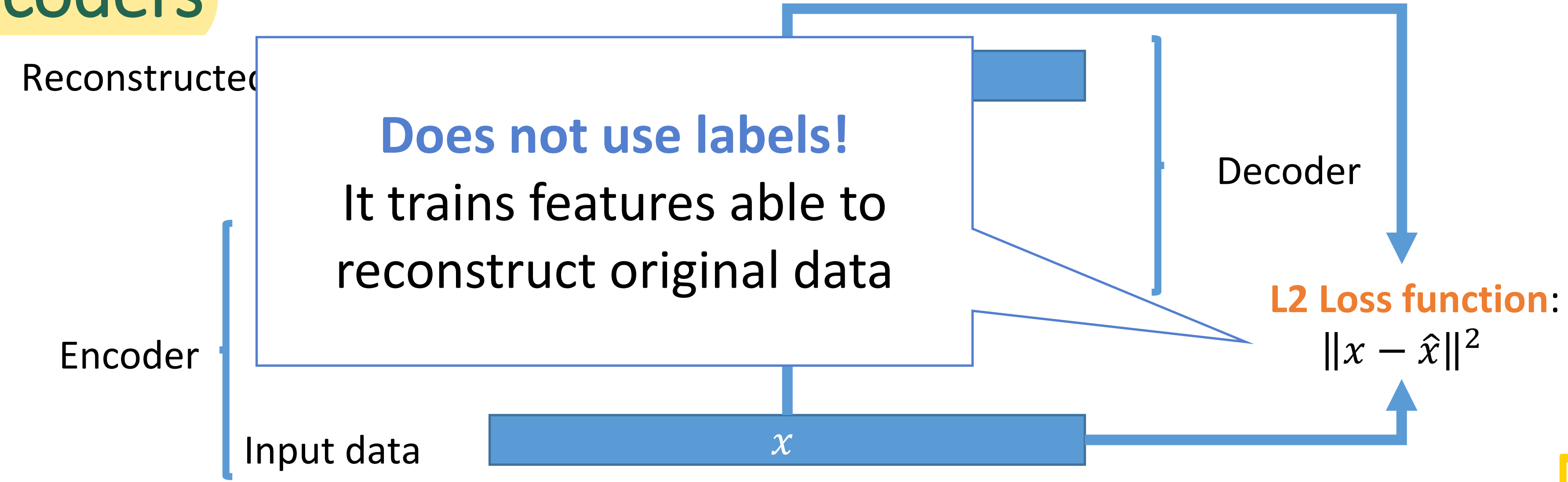
Input data

Encoder: 4-layer conv
Decoder: 4-layer upconv

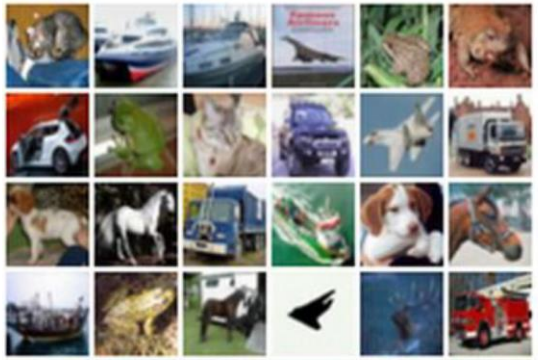
Reconstructed data



Autoencoders



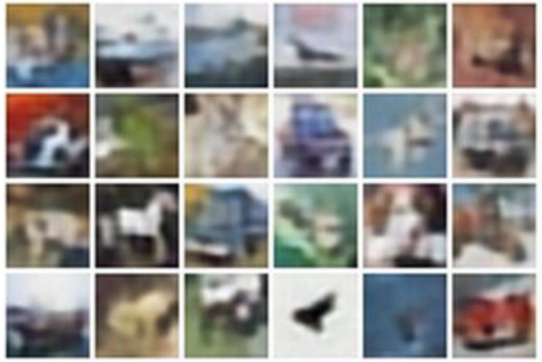
Does not use labels!
It trains features able to reconstruct original data



Input data

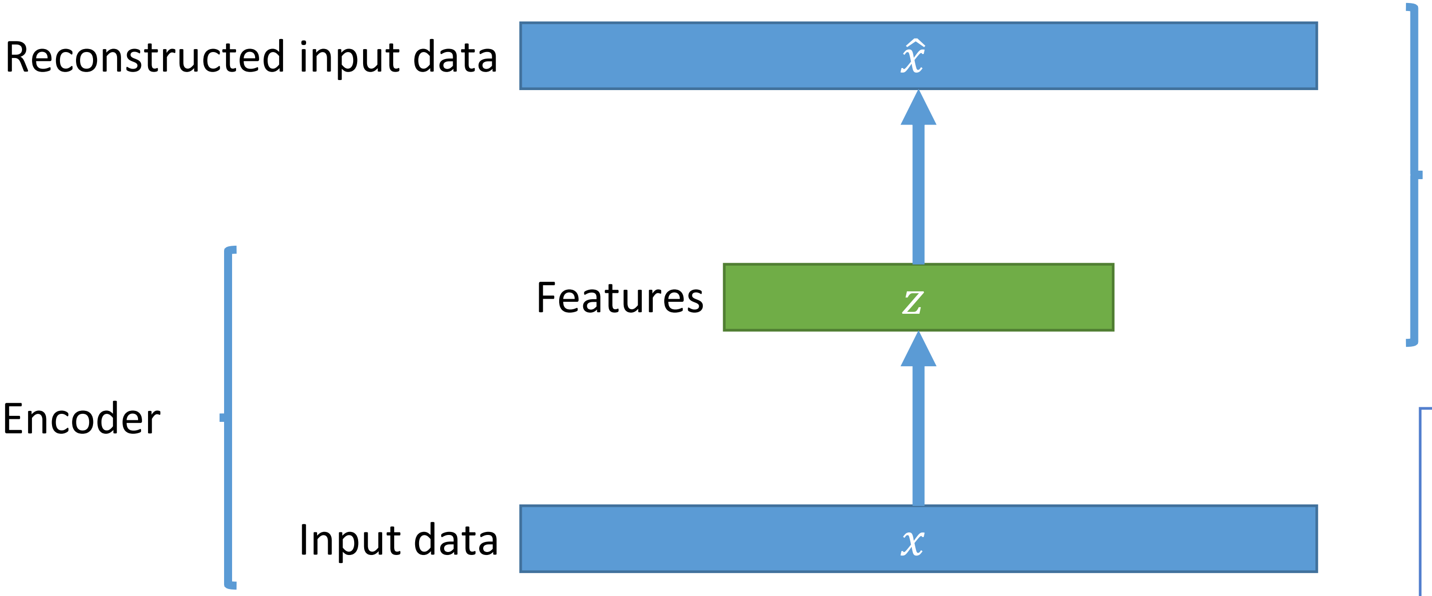
Encoder: 4-layer conv
Decoder: 4-layer upconv

Reconstructed data



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Autoencoders

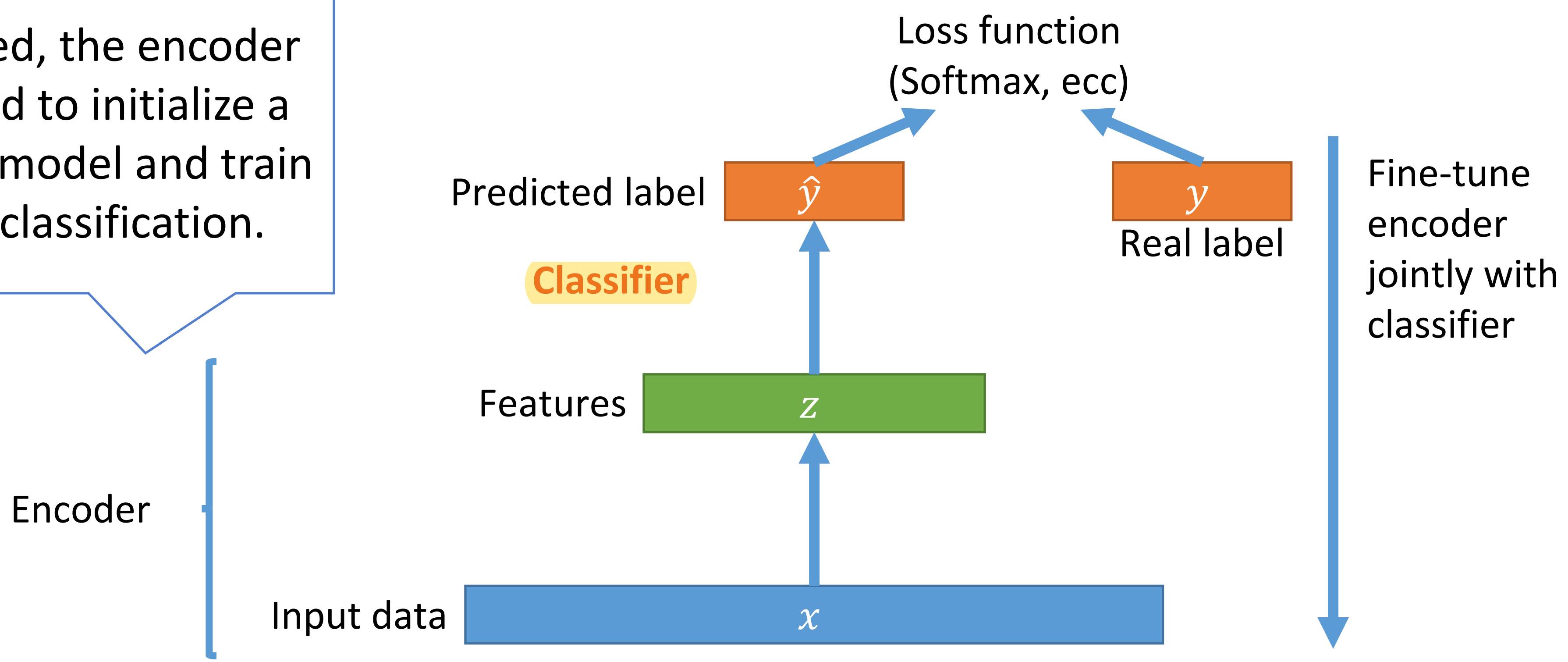


After training remove the decoder

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Autoencoders

Once trained, the encoder can be used to initialize a **supervised** model and train them for classification.

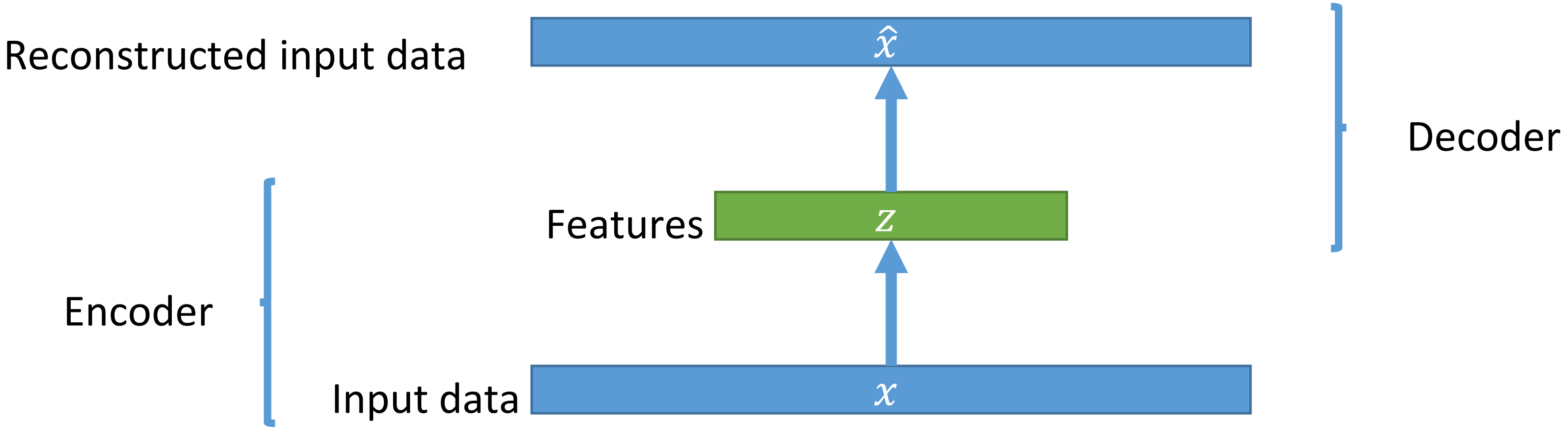


Autoencoders

Autoencoders can reconstruct data, and can learn features to initialize a supervised model

Features capture factors of variation in training data. Can we generate new images from an autoencoder?

Now that we have the model compiled and trained, we can generate new data, with the predict function of the model from samples of z (e.g. $z \sim \mathcal{N}(0, I)$).



Autoencoders

- Thus, we can use autoencoders to generate new data
  - Possibly, artificial training data
- Generated data will be similar to original training data but of poorer quality.
  - E.g., zooming an image or reducing the bit rate of a mp3 file
- There are many possibilities to improve the quality of the generated data
  - Variational Autoencoders
  - Boltzmann Machine
  - **GAN**

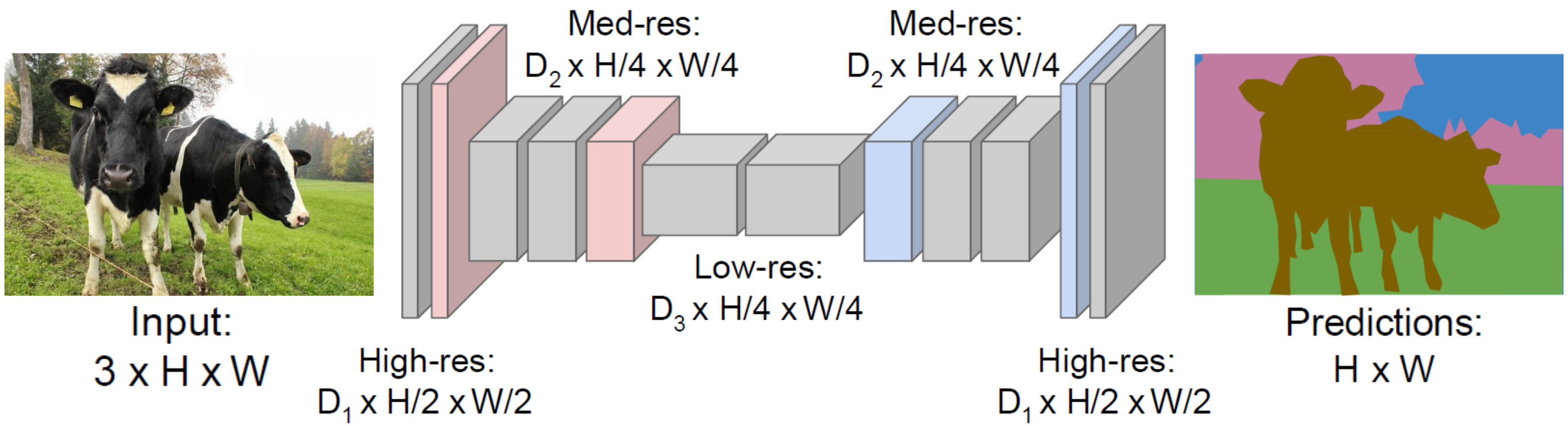
Regione Emilia-Romagna

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

Semantic Segmentation
Idea: Fully Convolutional

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!



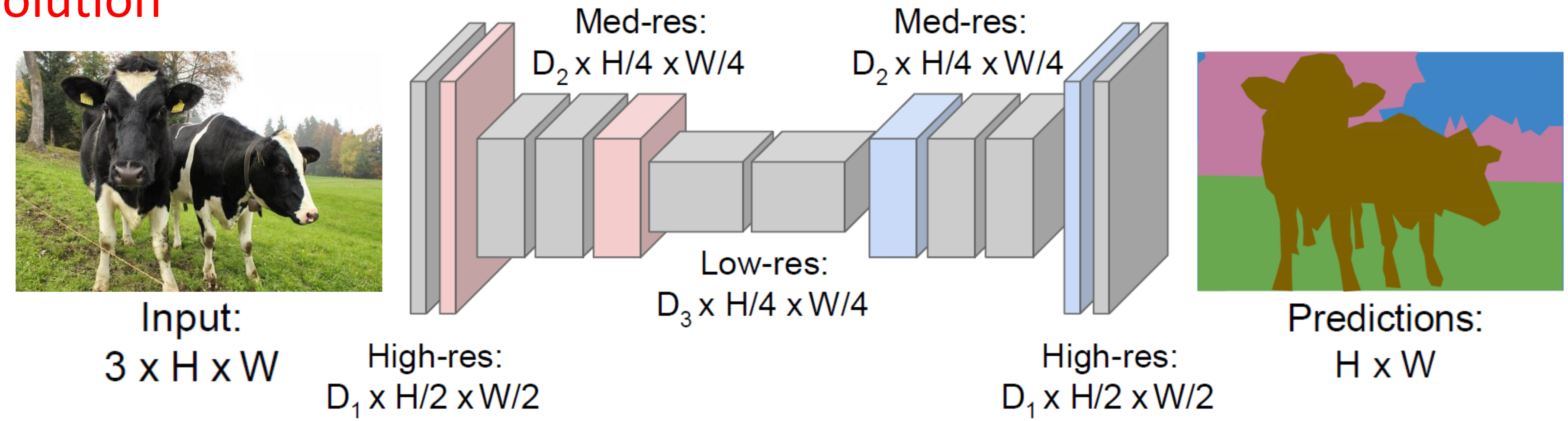
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Semantic Segmentation
Idea: Fully Convolutional

Downsampling
Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling
???



Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

In-network Upsampling
Unpooling

Nearest Neighbor

1	2
3	4

Input: 2 x 2

→

1	1	2	2
1	1	2	2
3	3	4	4
3	3	4	4

Output: 4 x 4

"Bed of Nails"

1	2
3	4

Input: 2 x 2

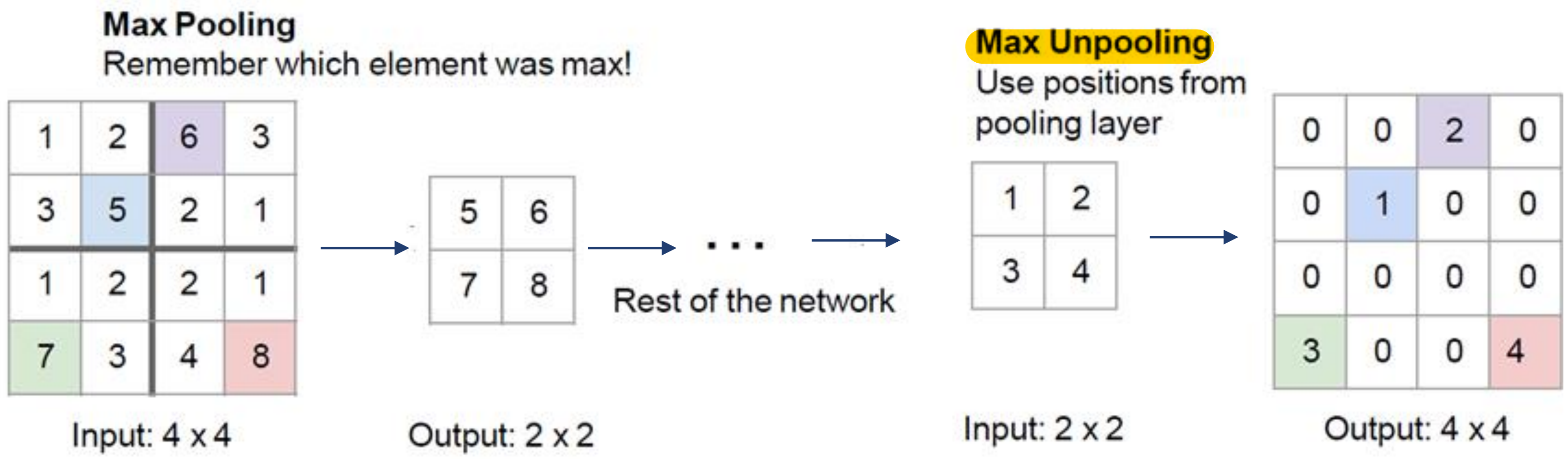
→

1	0	2	0
0	0	0	0
3	0	4	0
0	0	0	0

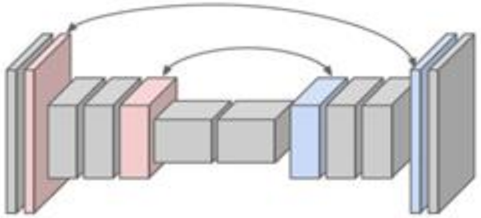
Output: 4 x 4

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

In-network Upsampling
Max Unpooling



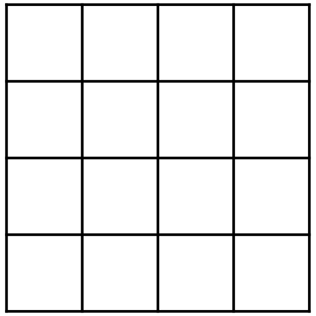
Corresponding pairs of downsampling and upsampling layers



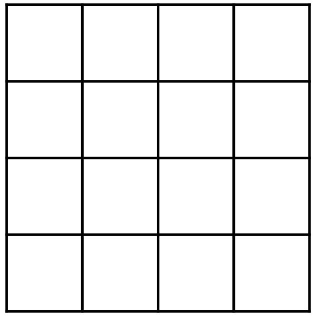
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Learnable Upsampling
Transpose Convolution

Recall: Typical 3 x 3 convolution, stride 1 pad 1



Input: 4x4



Output: 4x4

Unpooling are fixed functions, here we learn some weigths guiding the upsample

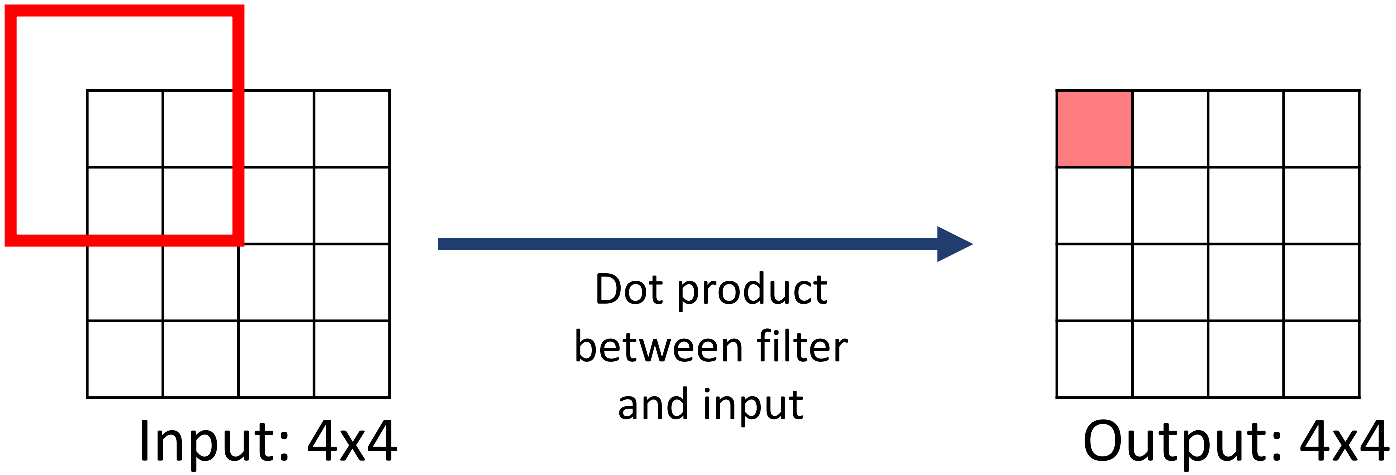
Regione Emilia-Romagna

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

Learnable Upsampling
Transpose Convolution

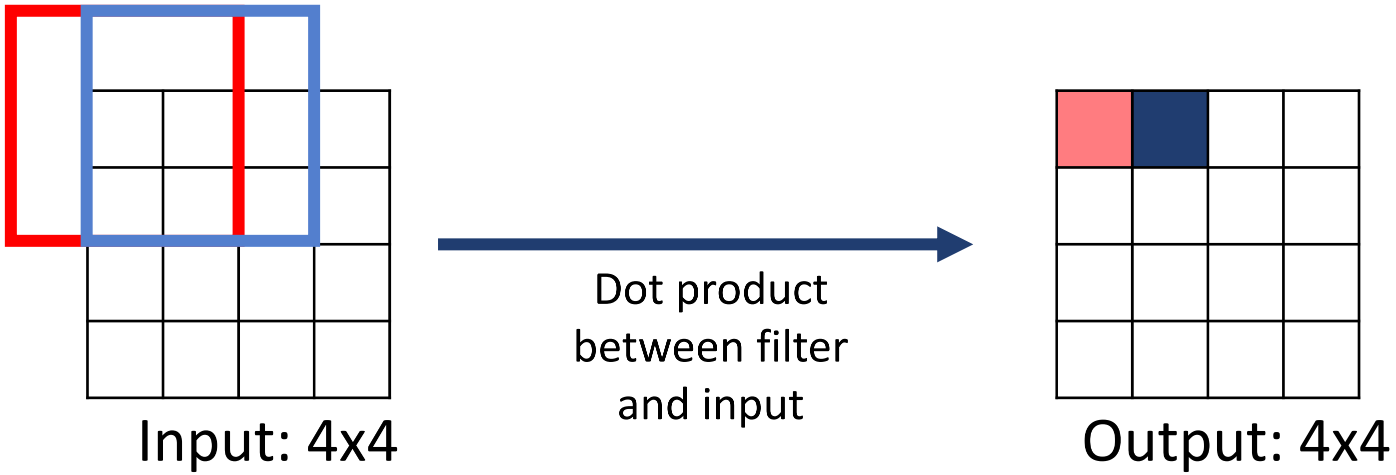
Recall: Typical 3 x 3 convolution, stride 1 pad 1



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Learnable Upsampling
Transpose Convolution

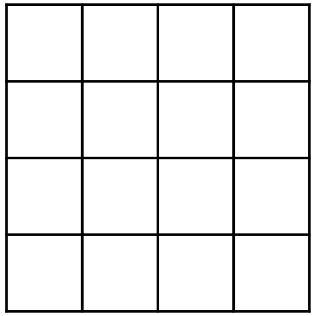
Recall: Typical 3 x 3 convolution, stride 1 pad 1



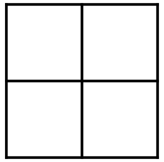
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Learnable Upsampling
Transpose Convolution

Recall: Typical 3 x 3 convolution, **stride 2** pad 1



Input: 4x4

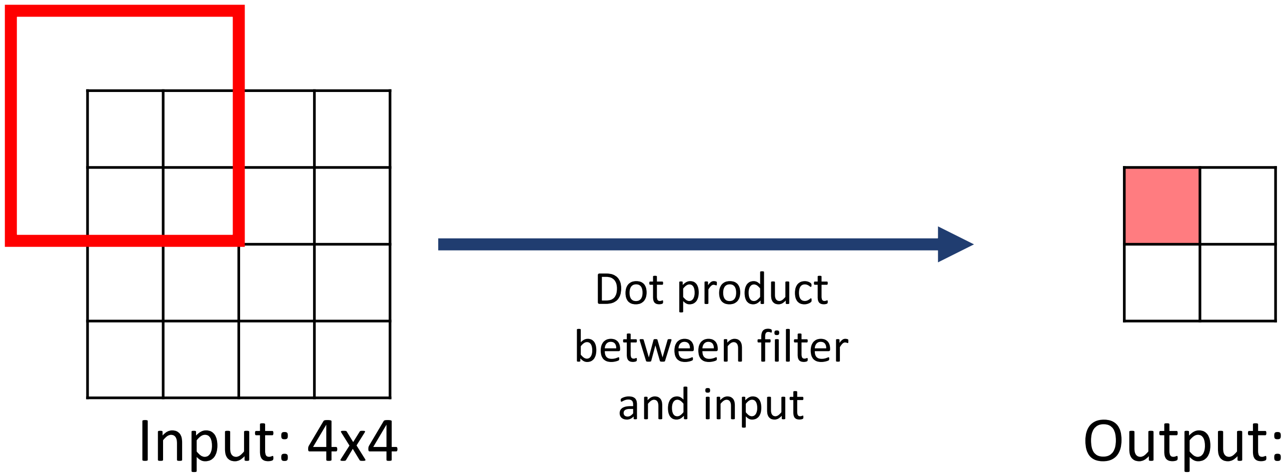


Output: 2x2

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

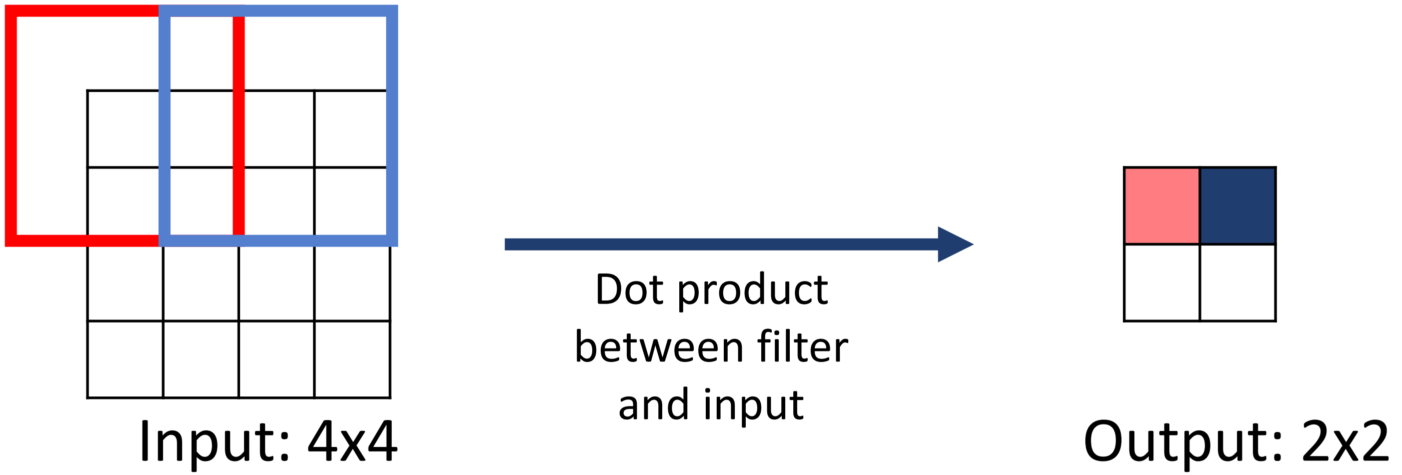
Learnable Upsampling
Transpose Convolution

Recall: Typical 3 x 3 convolution, **stride 2** pad 1



Learnable Upsampling
Transpose Convolution

Recall: Typical 3 x 3 convolution, **stride 2** pad 1

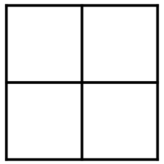


Filter moves 2 pixels in the input for every one pixel in the output

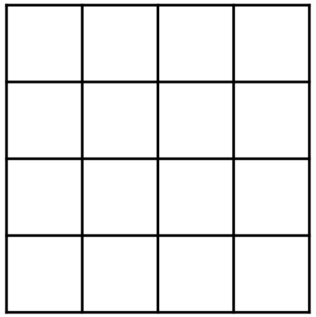
Stride gives ratio between movement in input and output

Learnable Upsampling
Transpose Convolution

3 x 3 **transpose** convolution, stride 2 pad 1



Input: 2x2

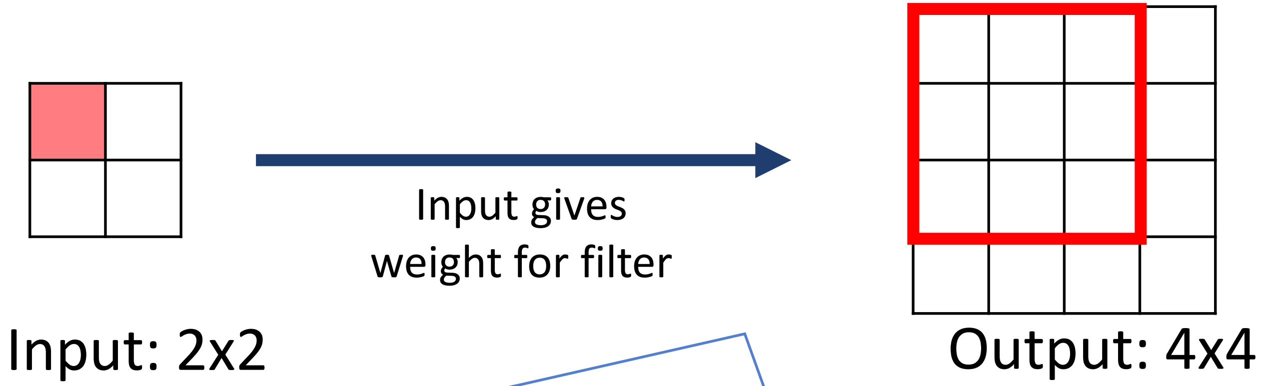


Output: 4x4

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Learnable Upsampling
Transpose Convolution

3 x 3 **transpose** convolution, stride 2 pad 1

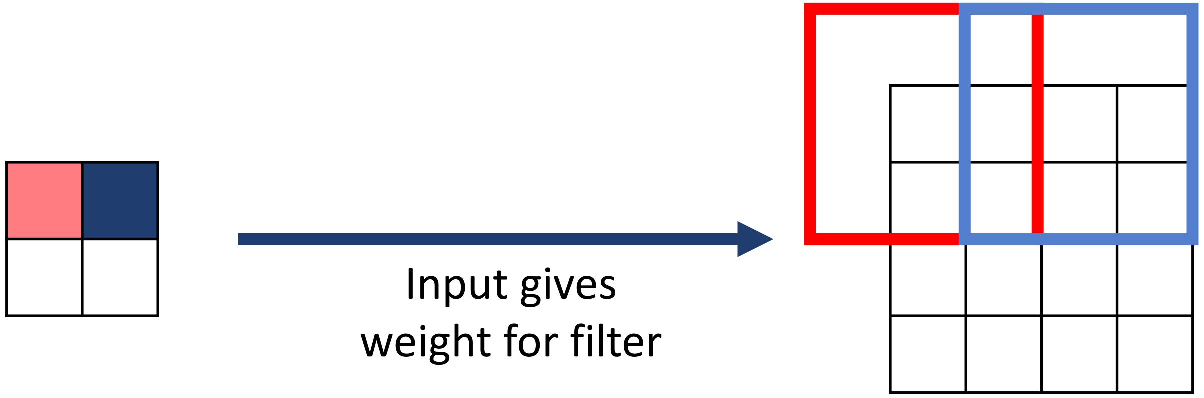


We take the value of the input, multiply the filter with the value and copy the result in the 3x3 zone in output.

Learnable Upsampling
Transpose Convolution

Sum where output overlaps

3 x 3 **transpose** convolution, stride 2 pad 1



Input: 2x2

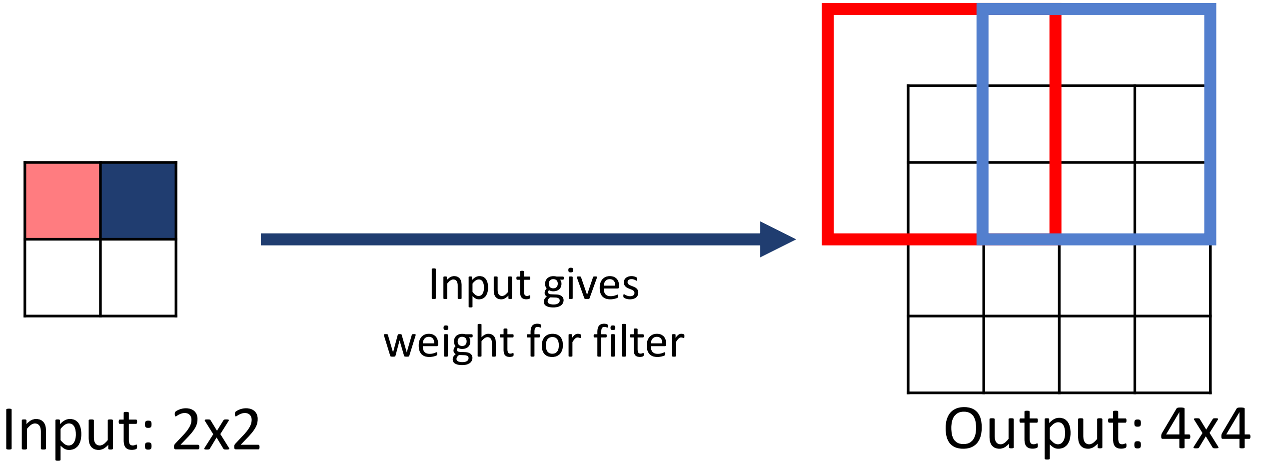
Output: 4x4

Filter moves 2 pixels in the **output** for every one pixel in the **input**

Stride gives ratio between movement in output and input

Learnable Upsampling
Transpose Convolution

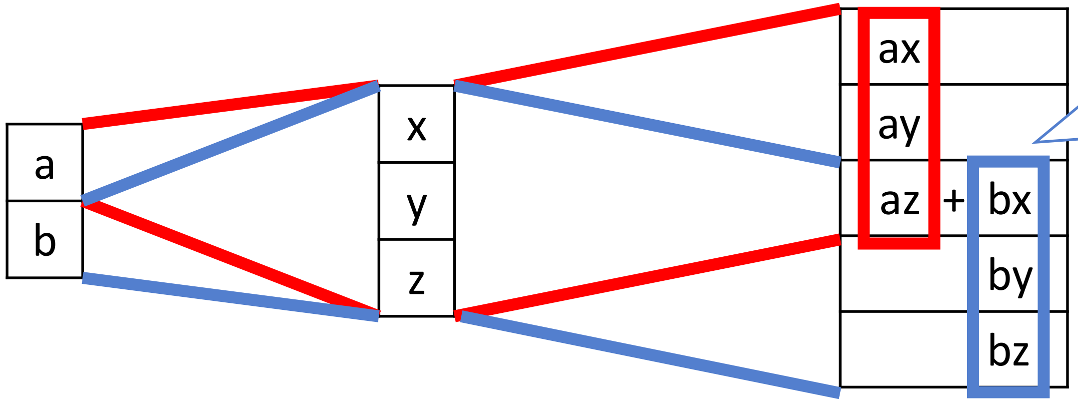
3 x 3 **transpose** convolution, stride 2 pad 1



Other names:

- Deconvolution
- Upconvolution
- Fractionally strided convolution
- Backward strided convolution

Transpose Convolution
1D Example



The stride is important here to avoid summing to much values for each shared cell.

Output contains copies of the filter weighted by the input, summing at where it overlaps in the output

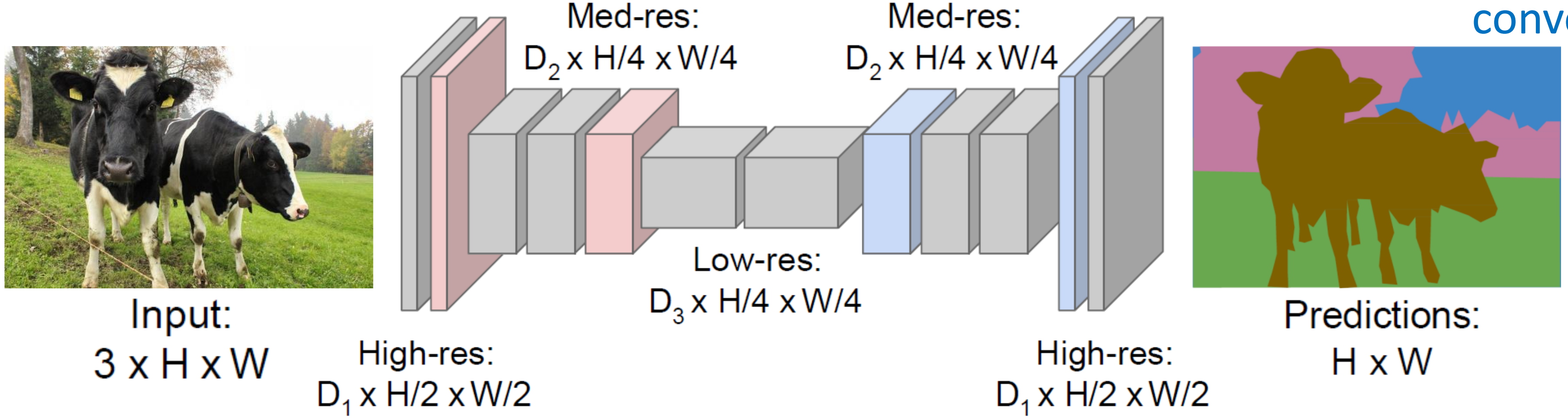
Need to crop one pixel from output to make output exactly 2x input

Semantic Segmentation Idea: Fully Convolutional

Downsampling
Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling
Unpooling or strided transpose convolution



Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Autoencoders

- Thus, we can use autoencoders to generate new data
  - Possibly, **new artificial training data**
- Generated data will be similar to original training data but of poorer quality.
  - E.g., zooming an image or reducing the bit rate of a mp3 file
- There are many possibilities to improve the quality of the generated data
  - Variational Autoencoders
  - Boltzmann Machine
  - **GAN**

Generative Adversarial Networks

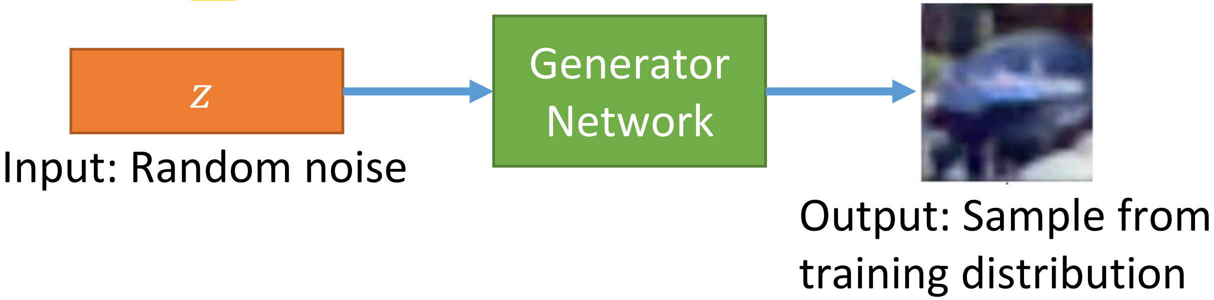
- The main difference with respect Autoencoders is that instead of learn and model a density, it just samples data miming a distribution without explicitly model it
- GANs don’t work with any explicit density function!
- Instead, take game-theoretic approach: learn to generate from training distribution through 2-player game

Generative Adversarial Networks

- Learn to generate from training distribution through 2-player game
  - **Generator network**: try to fool the discriminator by generating real-looking images
  - **Discriminator network**: try to distinguish between real and fake images

Generator Network

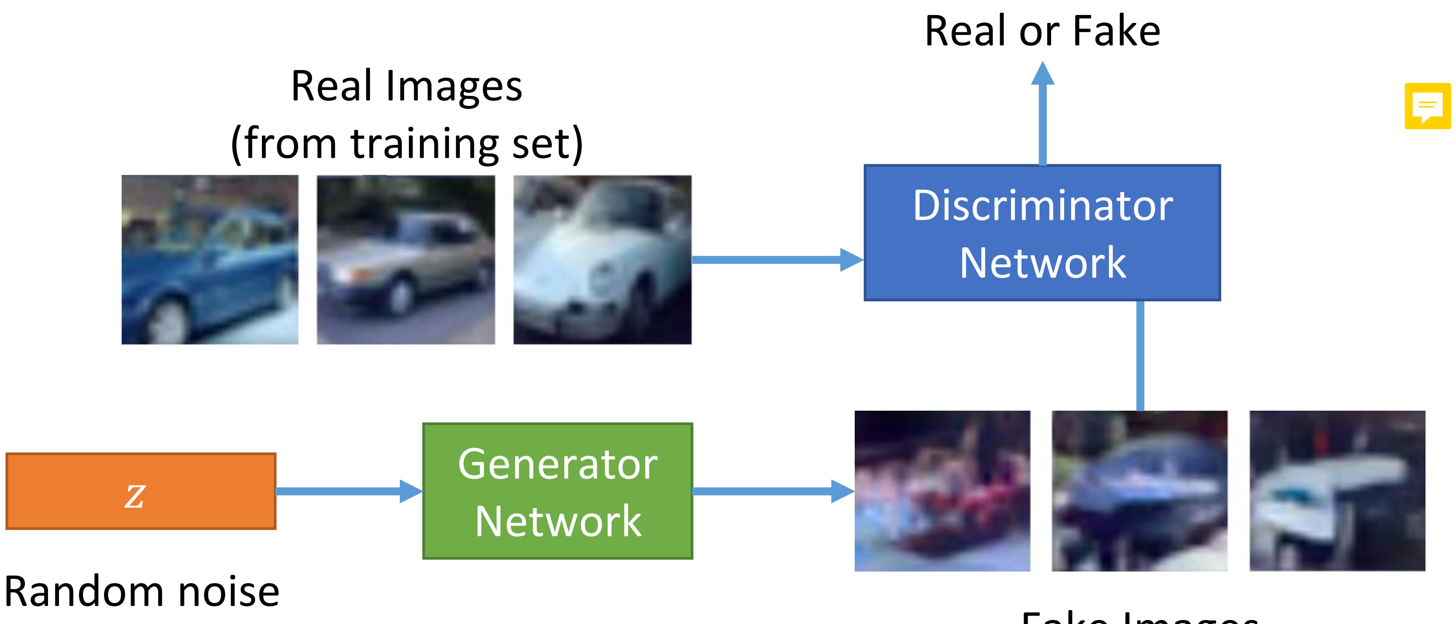
- Problem: Want to sample from complex, high-dimensional training distribution. No direct way to do this!
- Solution: Sample from a simple distribution, e.g. random noise. Learn transformation to training distribution using NNs



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Generative Adversarial Networks

- **Generator network**: try to fool the discriminator by generating real-looking images
- **Discriminator network**: try to distinguish between real and fake images



Ian Goodfellow et al., “Generative Adversarial Nets”, NIPS 2014
Fake and real images copyright Emily Denton et al. 2015. Reproduced with permission.

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Train GANs: Two-Player game

- **Generator network**: try to fool the discriminator by generating real-looking images
- **Discriminator network**: try to distinguish between real and fake images

→ Train jointly in **minimax game**

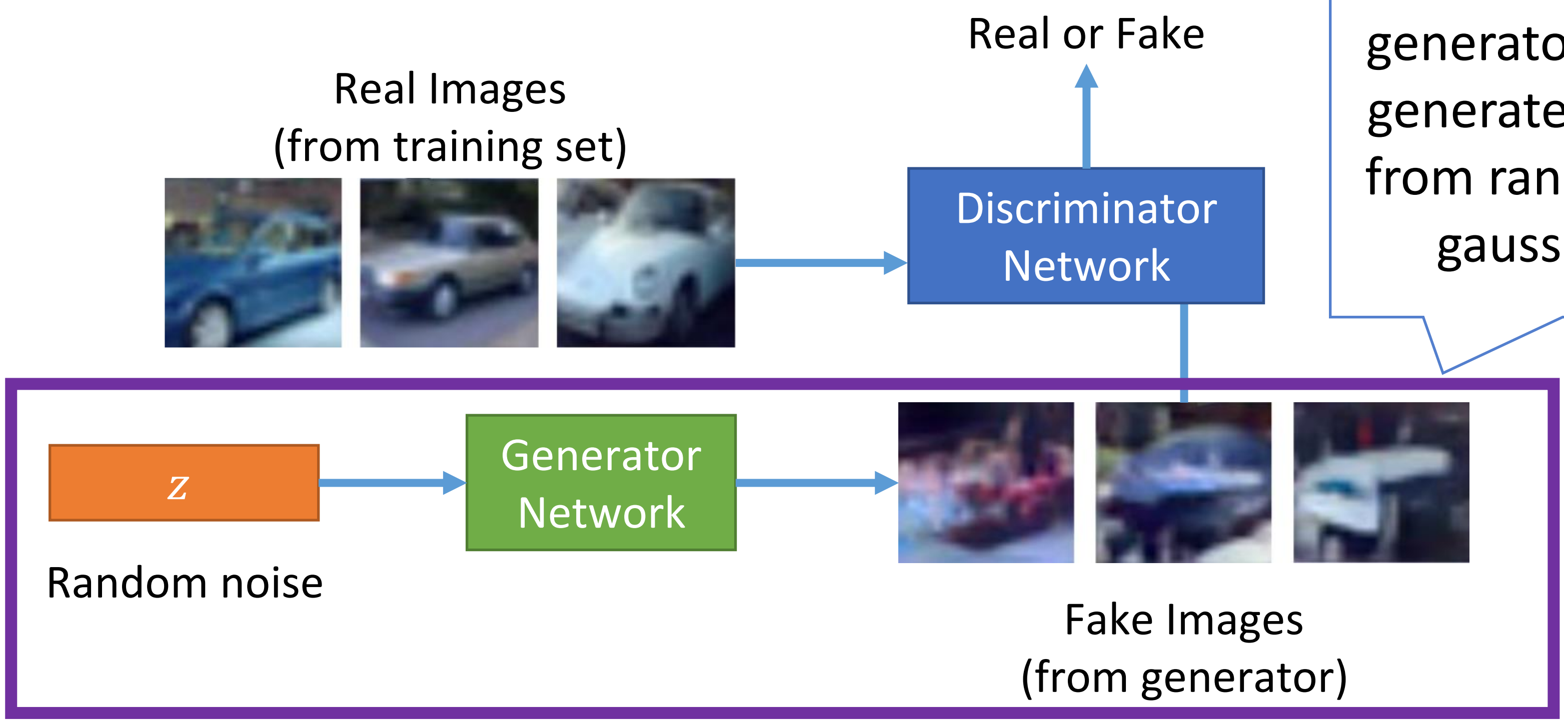
- Training is done by gradient **ascent**, which facilitates the training more than gradient descent. However jointly training two networks is challenging, can be unstable.
  - Refer to Ian Goodfellow et al., “Generative Adversarial Nets”, NIPS 2014 for detail on the training.

- This is an active area of research.

Generative Adversarial Networks

- **Generator network**: try to fool the discriminator by generating real-looking images
- **Discriminator network**: try to distinguish between real and fake images

After training, use generator network to generate new images from random (usually gaussian) noise



Ian Goodfellow et al., "Generative Adversarial Nets", NIPS 2014
Fake and real images copyright Emily Denton et al. 2015. Reproduced with permission.