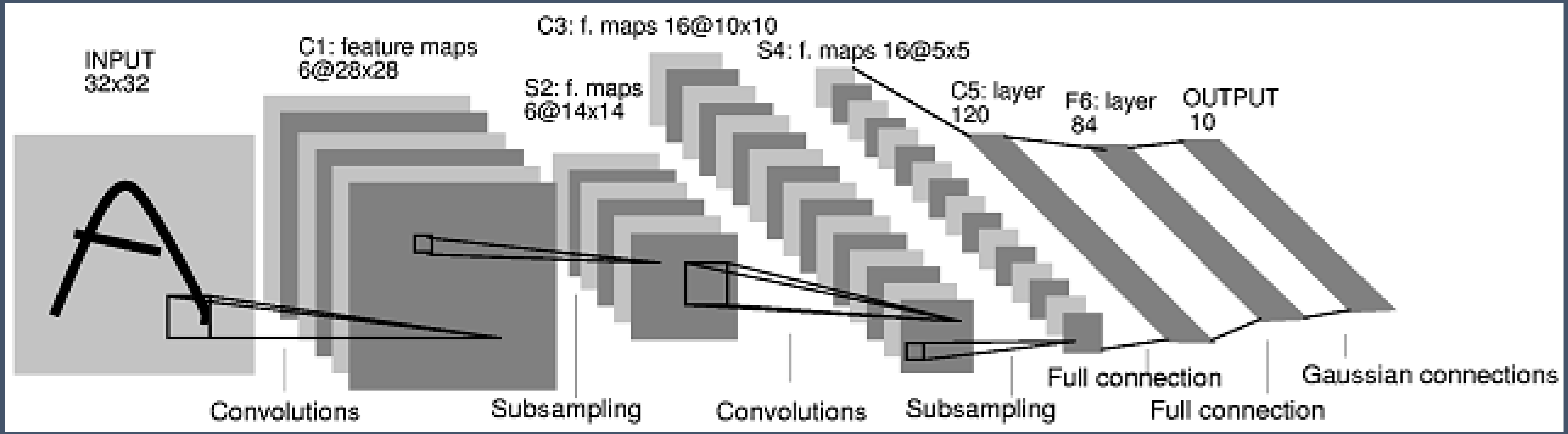
CNNs
and
Applications for medical imaging

TECHNION
Israel Institute
of Technology

Short review of CNN architectures

- CNN Evolution
  - Classification
    - LeNet
    - AlexNet
    - VGG
    - GoogLeNet / Inception arch.
    - ResNet
  - Segmentation
    - FCN
    - U-Net
- What to do when there is little data
  - Augmentation
  - Transfer learning
  - Auto-encoders (Unsupervised Learning)
  - Semi-supervised learning
  - GANs

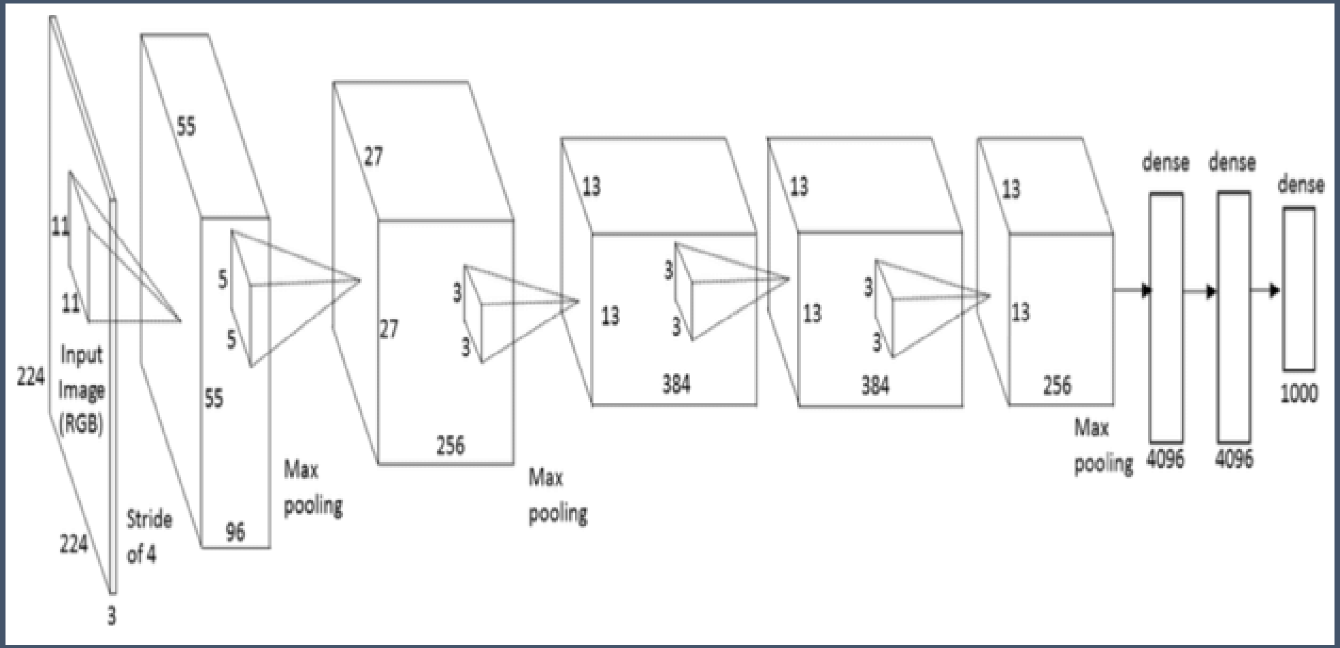
Le Net
1998



- One of the first NN first’s (best MNIST)
- Considered the first reference CNN architecture
- Today, a relatively shallow network

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, November 1998

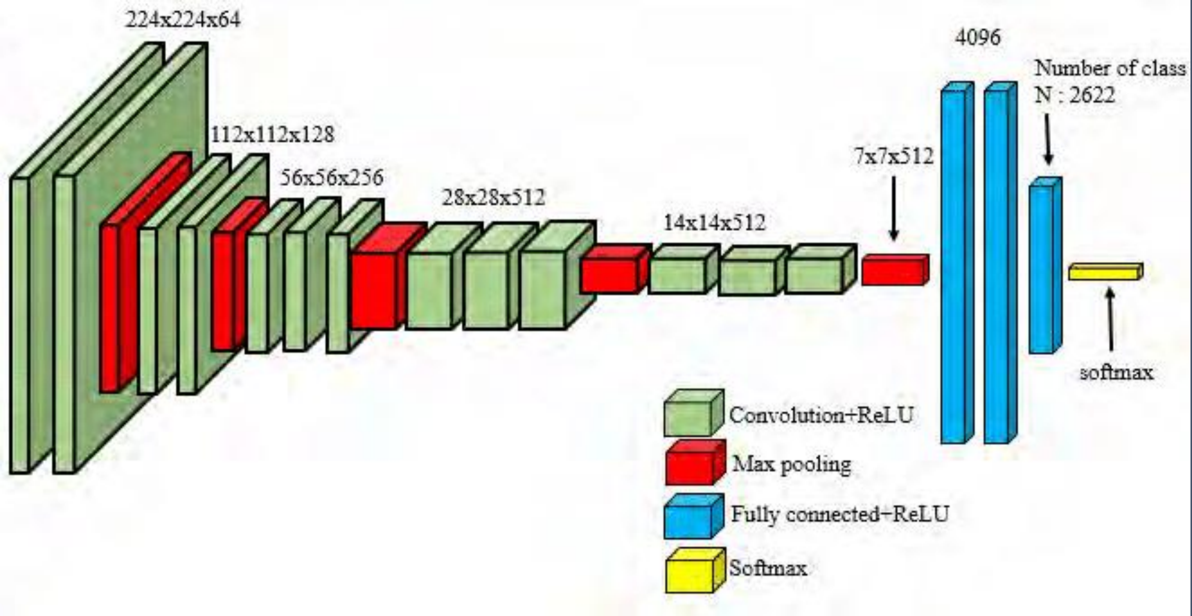
Alex net
2012



- Won the 2012 ILSVRC (15.3% error for top-5) with a huge margin >10%
- Programed on GPU
- Opened the current DL era

A. Krizhevsky, G. E. Hinton, ImageNet classification with deep convolutional neural networks.
In NIPS, pp. 1106–1114, 2012

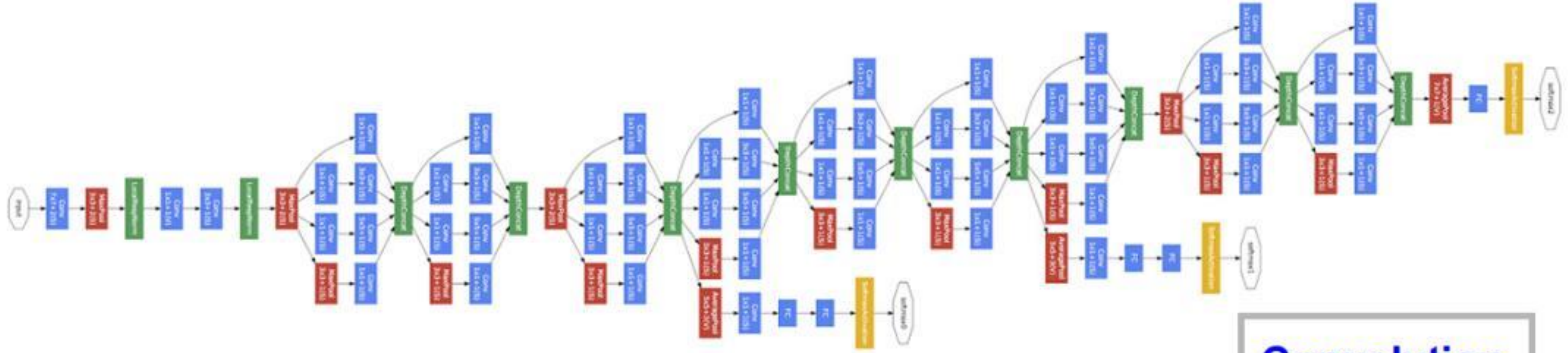
VGG
2014



- 2^{nd} @ 2014 ILSVRC
- Possibly the most popular architecture
- Modules: (Pool + (2 or 3)x$\text{Conv}_{(\text{x2features})}$)

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015

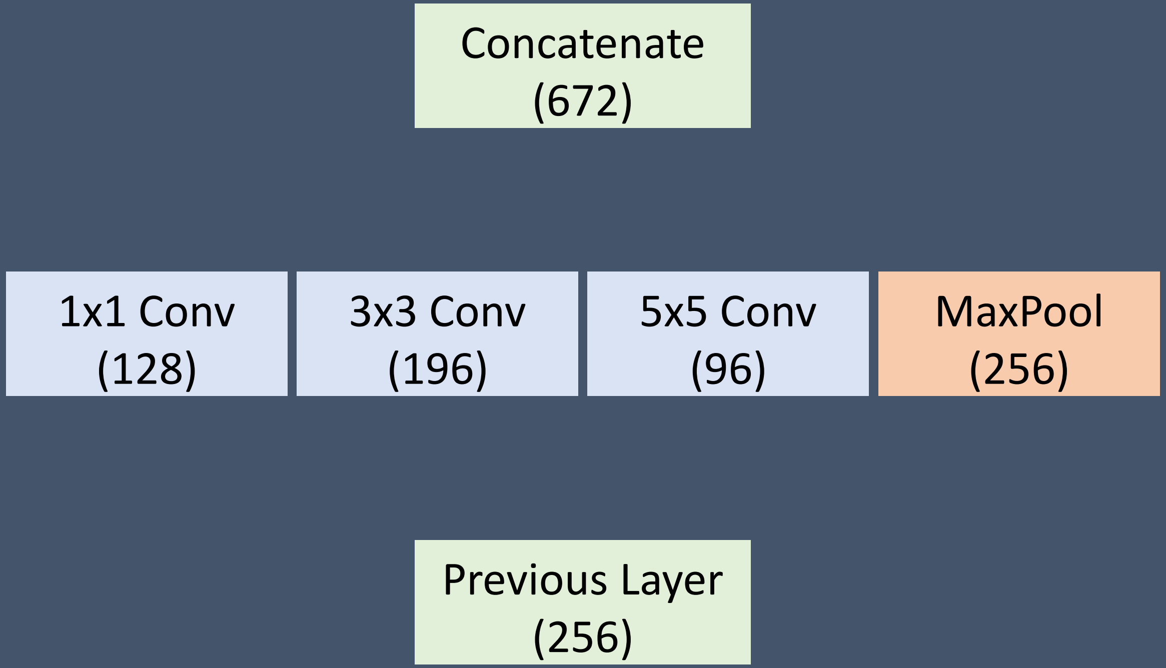
GoogLeNet, the Inception architecture
2014

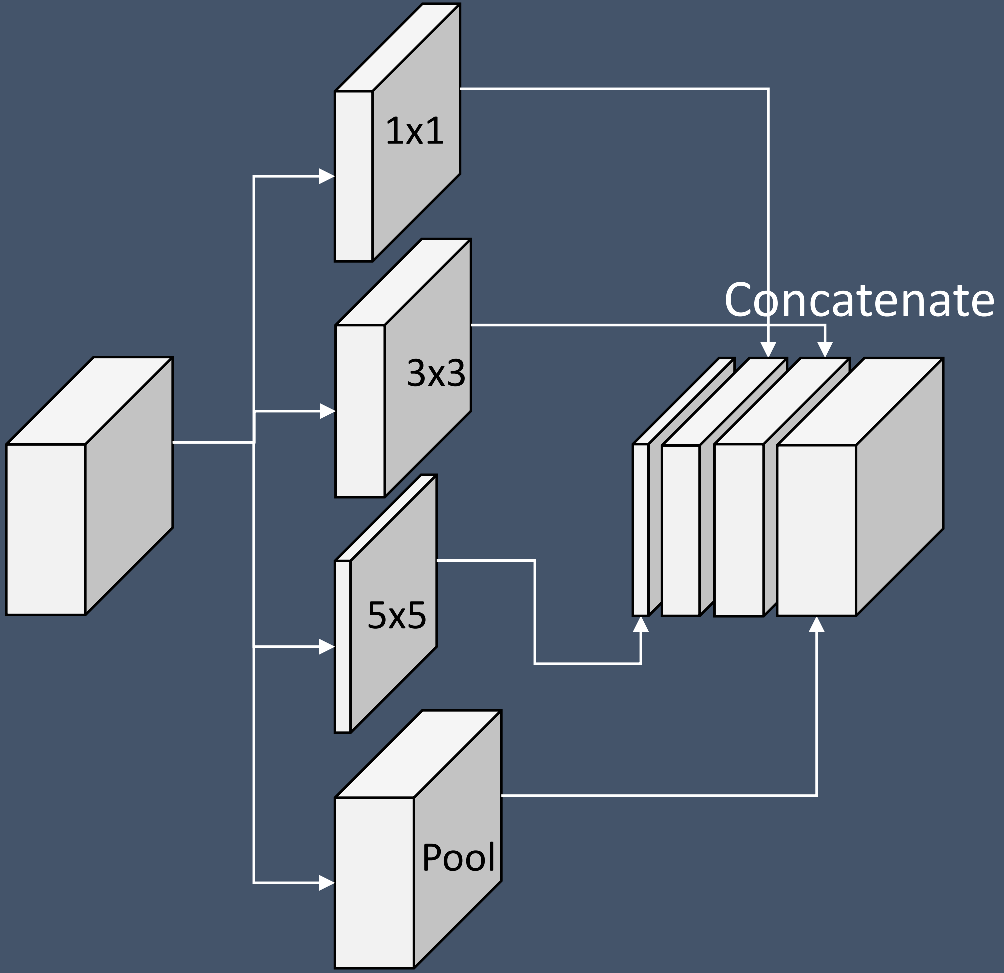


- Won the 2014 ILSVRC
- Modules: 9xInception module
- More nonlinearities, fewer parameters, less computation
- Considered 22 layers deep (max #nonlinearities in path)

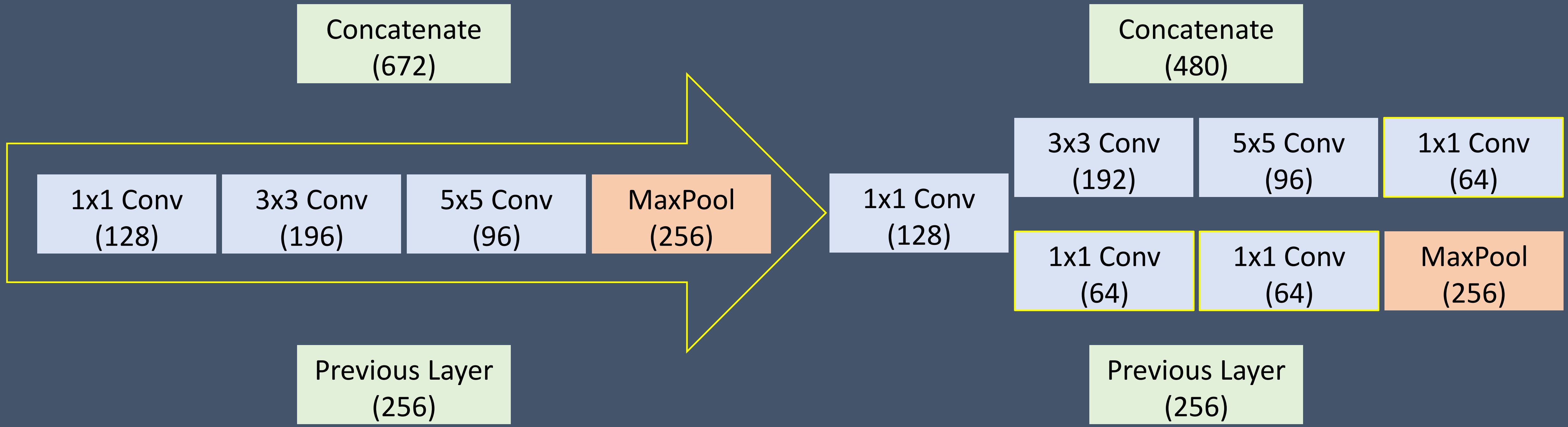
C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions. In CVPR, 2015.

Inception Module
Naïve

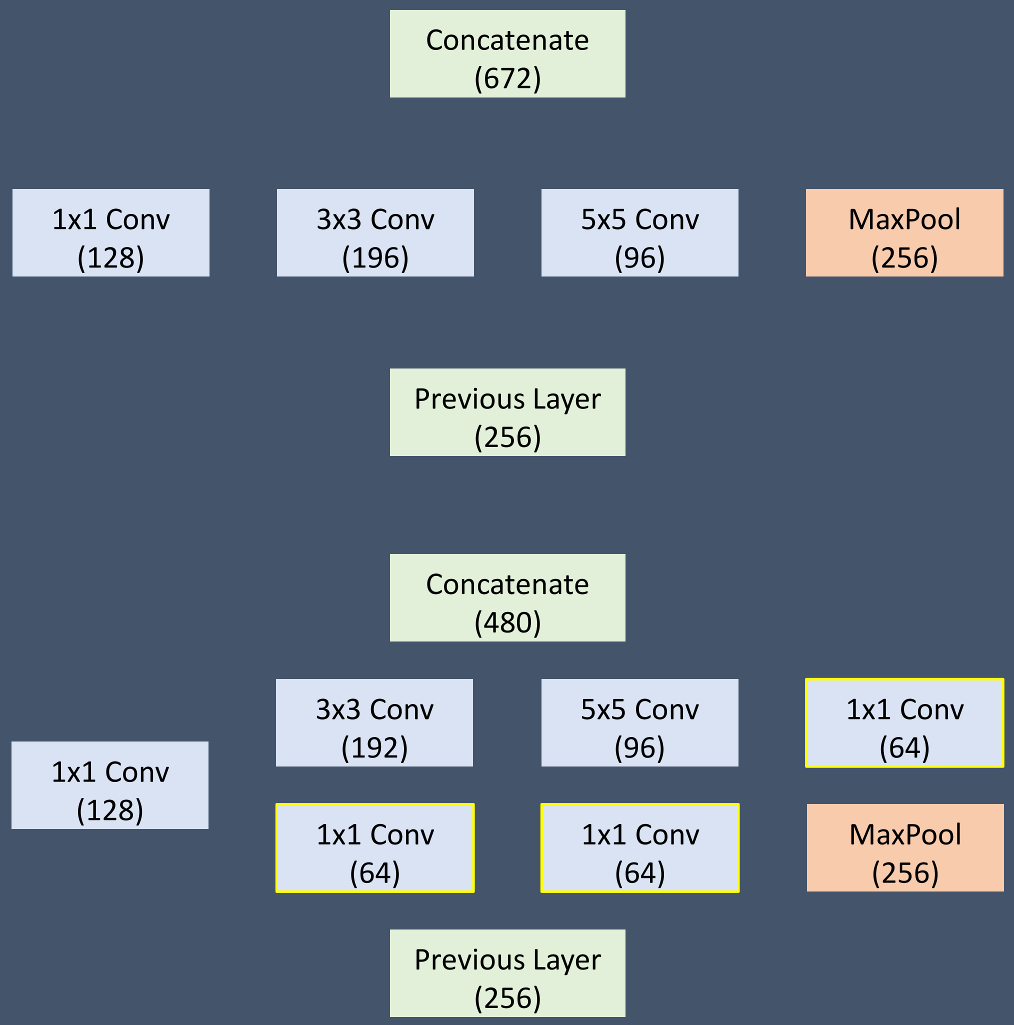




Inception Module
Implementation with bottlenecks



Inception Module
Reduced Computations

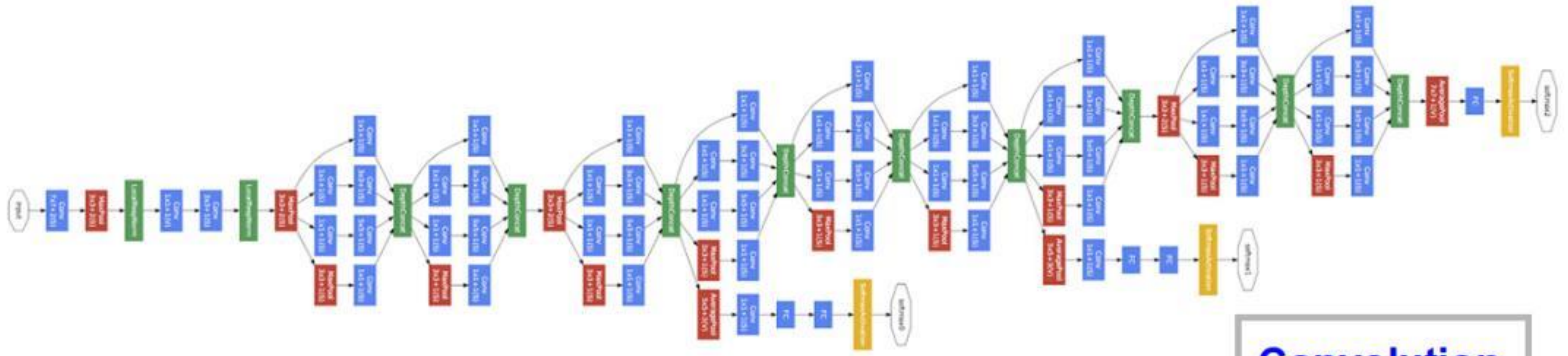


	Parameters (no biases)
1x1 conv	256x1x1x128=32.7K
3x3 conv	256x3x3x192=442K
5x5 conv	256x5x5x96=614.4K
Totals	1,089,536

1x1 conv	256x1x1x128=32.7K
1x1 conv(3)	256x1x1x64=16.4K
3x3 conv	64x3x3x192=110.6K
1x1 conv(5)	256x1x1x64=16.4K
5x5 conv	64x5x5x96=153.6.4K
1x1 conv(P)	256x1x1x64=16.4K
Totals	346,112

GoogleNet

Inception architecture details

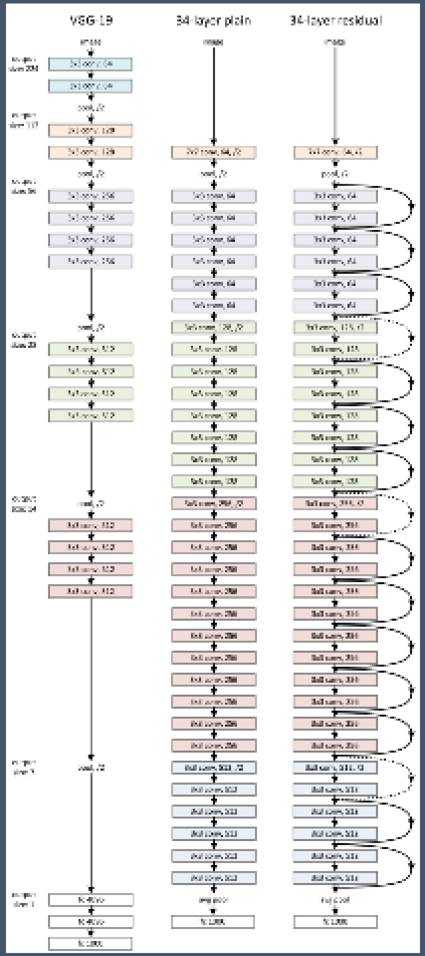


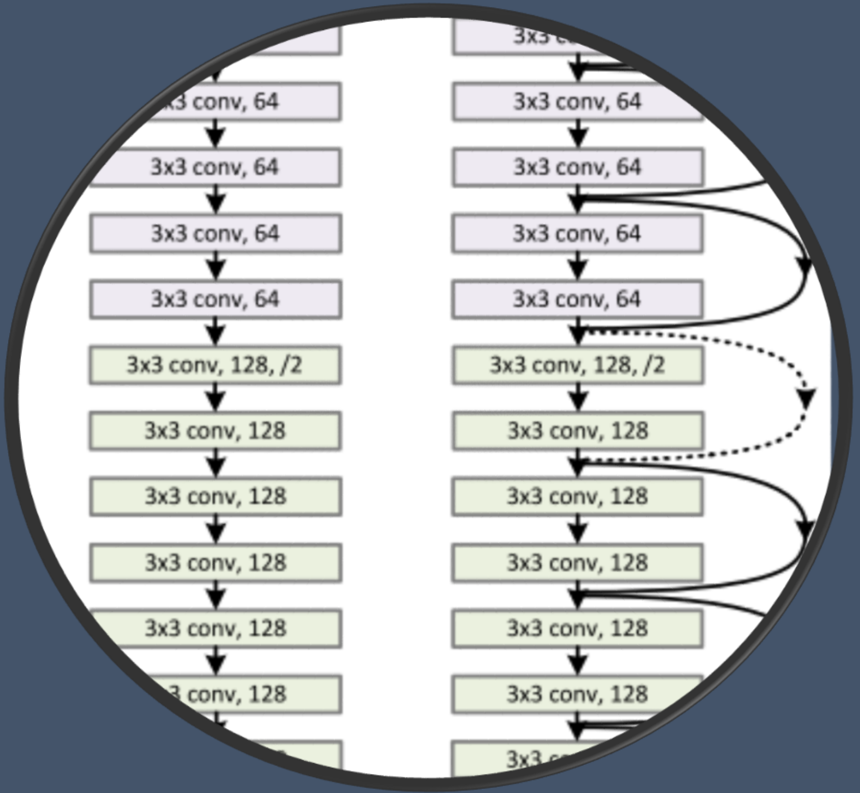
- Two auxiliary classification paths for loss injection
  - Overcoming vanishing gradients in deep layers
  - During training only
- No dense layers for classification
  - Pooling reduces size to ~1000

ResNet
2015

- Won the 2015 ILSVRC with top 5 error: 3.57%
- Broke the human benchmark of 5.1% error
- Modules: ResNet
- Going deeper: Many more nonlinearities

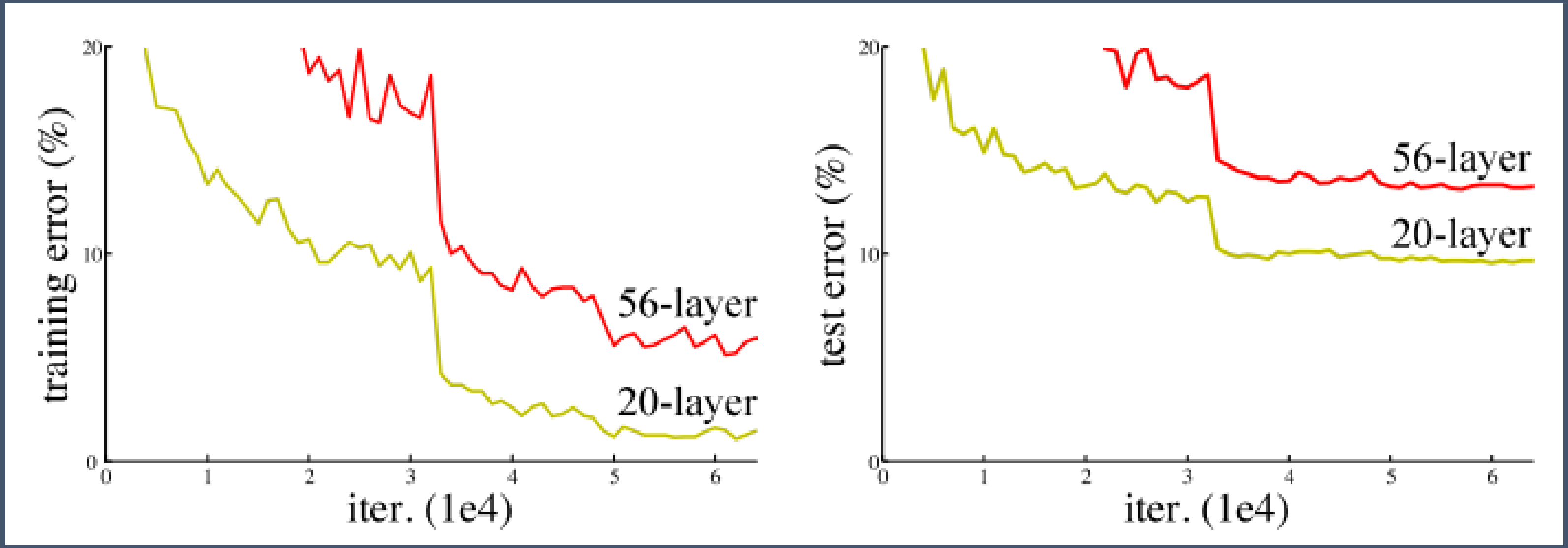
K. He X. Zhang S. Ren and J. Sun,
Deep Residual Learning for Image Recognition,
CVPR 2016





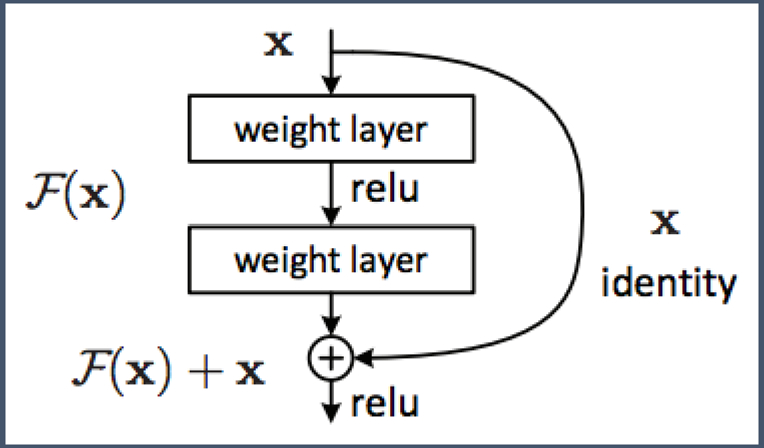
ResNet

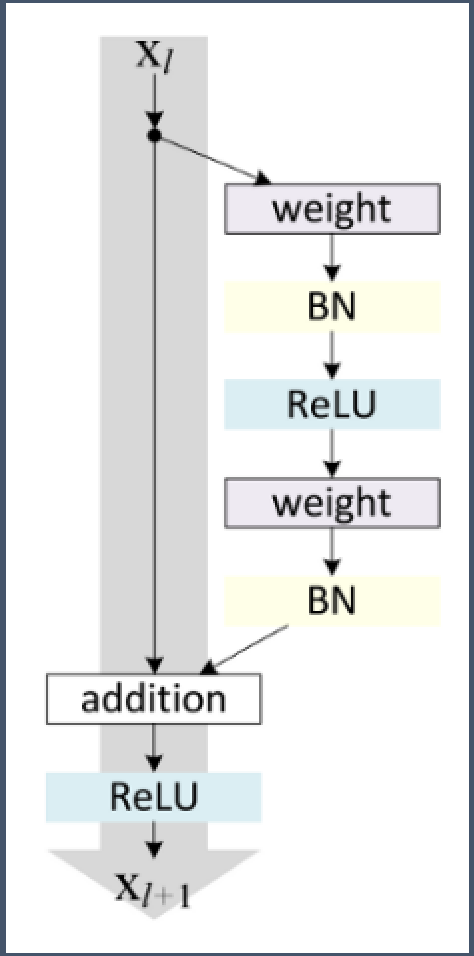
Motivation: Going deeper



- Deeper networks perform worse, but not due to overfitting
- This is strange
- There exists a better solution:
  - Deep layers = identity

ResNet
ResNet module



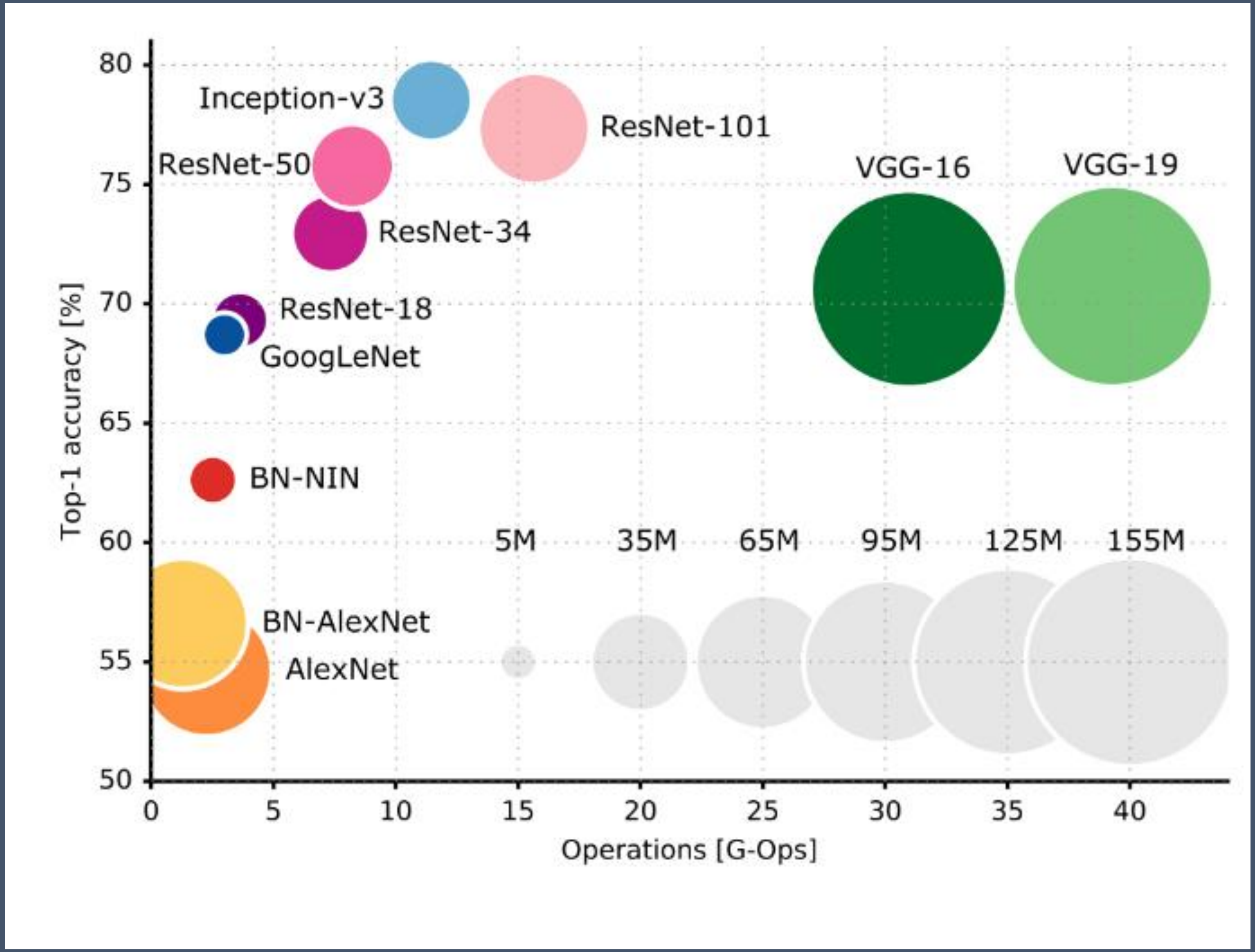


- ResNet modules compute residual signals $F(x)$
- Addition shortcuts the loss across blocks

- There are many variant details for the block
  - Convolution blocks are broken to their component layers

Short review of CNN architectures

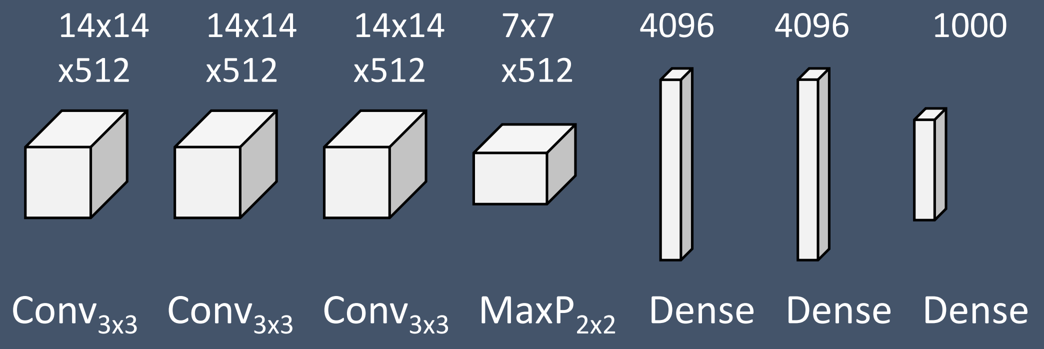
Comparing CNN architectures



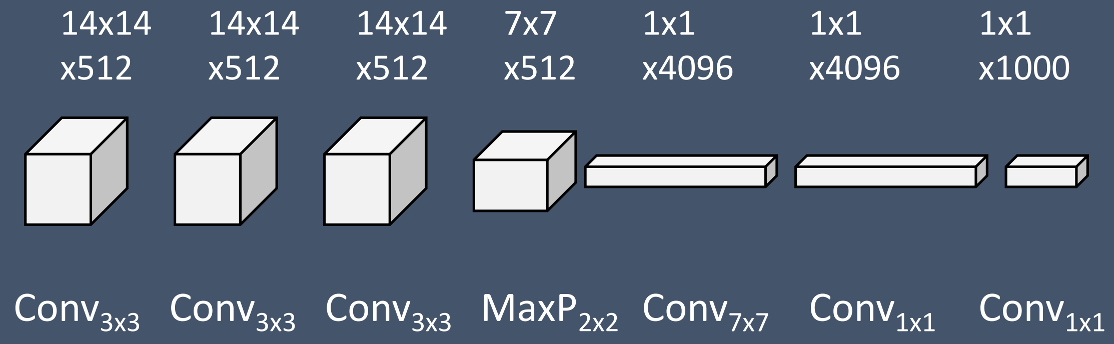
Fully convolutional architectures

1x1 Convolution kernels

Tail end of the VGG architecture




```
…
# Block 5
x = Conv2D(512, (3, 3), …)(x)
x = Conv2D(512, (3, 3), …)(x)
x = Conv2D(512, (3, 3), ...)(x)
x = MaxPooling2D((2, 2), strides=(2, 2))(x)
# Classification block
x = Flatten()(x)
x = Dense(4096, activation='relu')(x)
x = Dense(4096, activation='relu')(x)
x = Dense(classes, activation='softmax')(x)
```

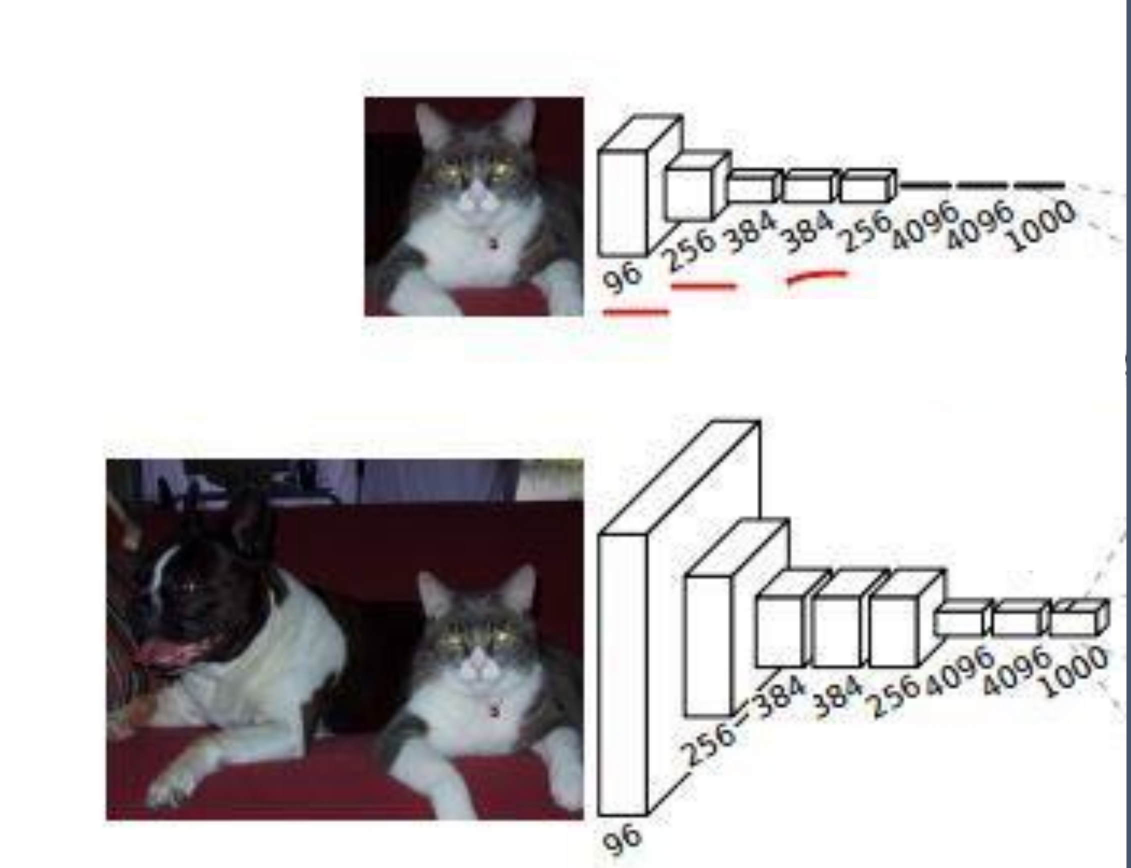



```
…
# Classification block
x = Conv2D(4096, (7, 7), padding=‘valid’…)(x)
x = Conve2D(4096, (1, 1), … )(x)
x = Conve2D(classes, (1, 1), … )(x)
```

Fully convolutional network

Arbitrary input size

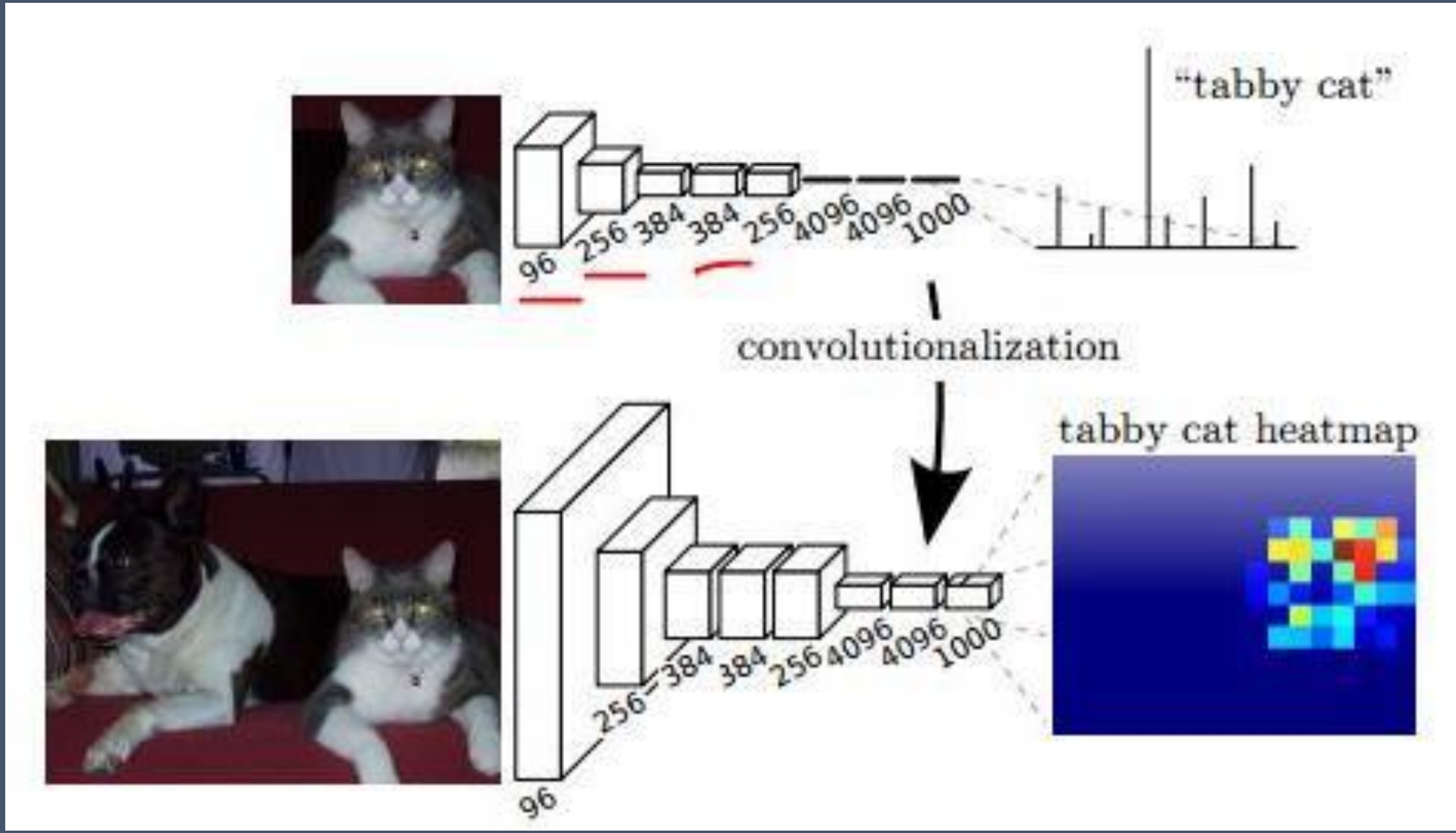
- Train on a nominal size
  - e.g. 224x224
- Infer on an arbitrary (larger) size
  - Results resized accordingly
  - To get a single class – average
- Spatial classification constitutes rough localization



J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation. CoRR, 2014.

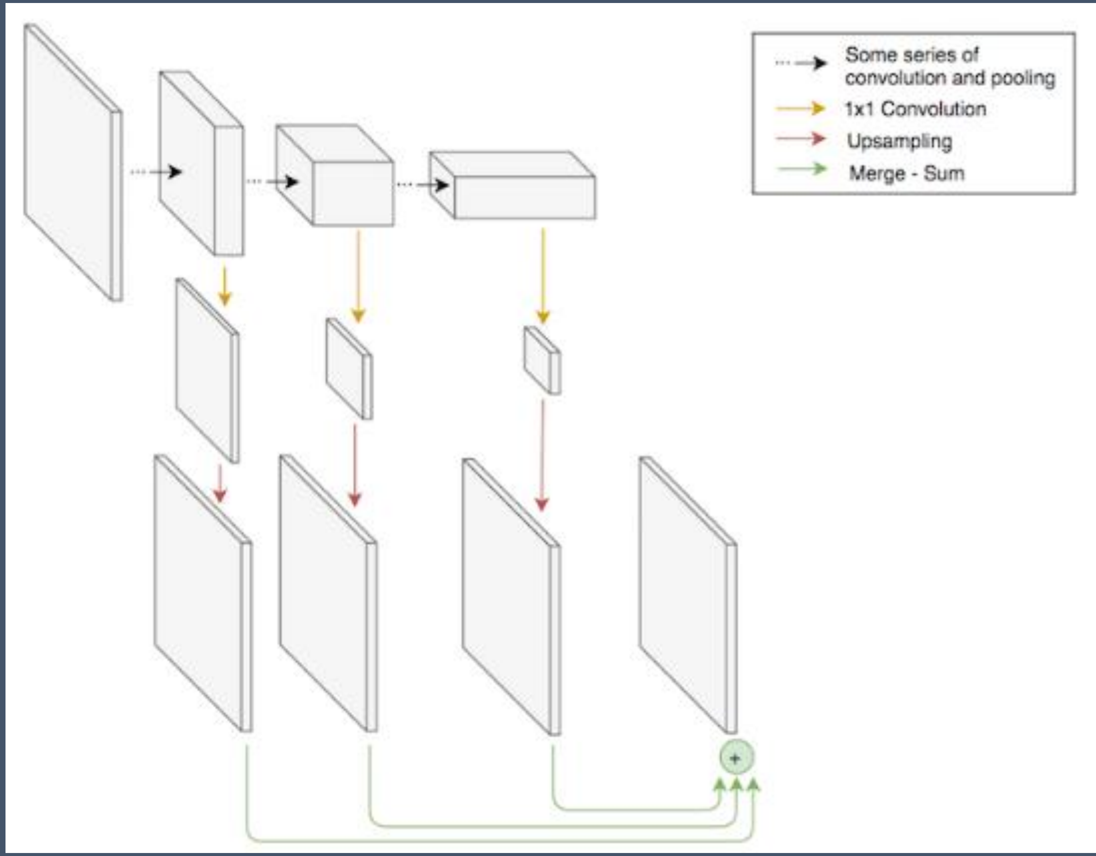
Fully convolutional architectures
Rough localization

- Localization for free
- Fine-grained details lost in consecutive down-sampling
- Can we do better?



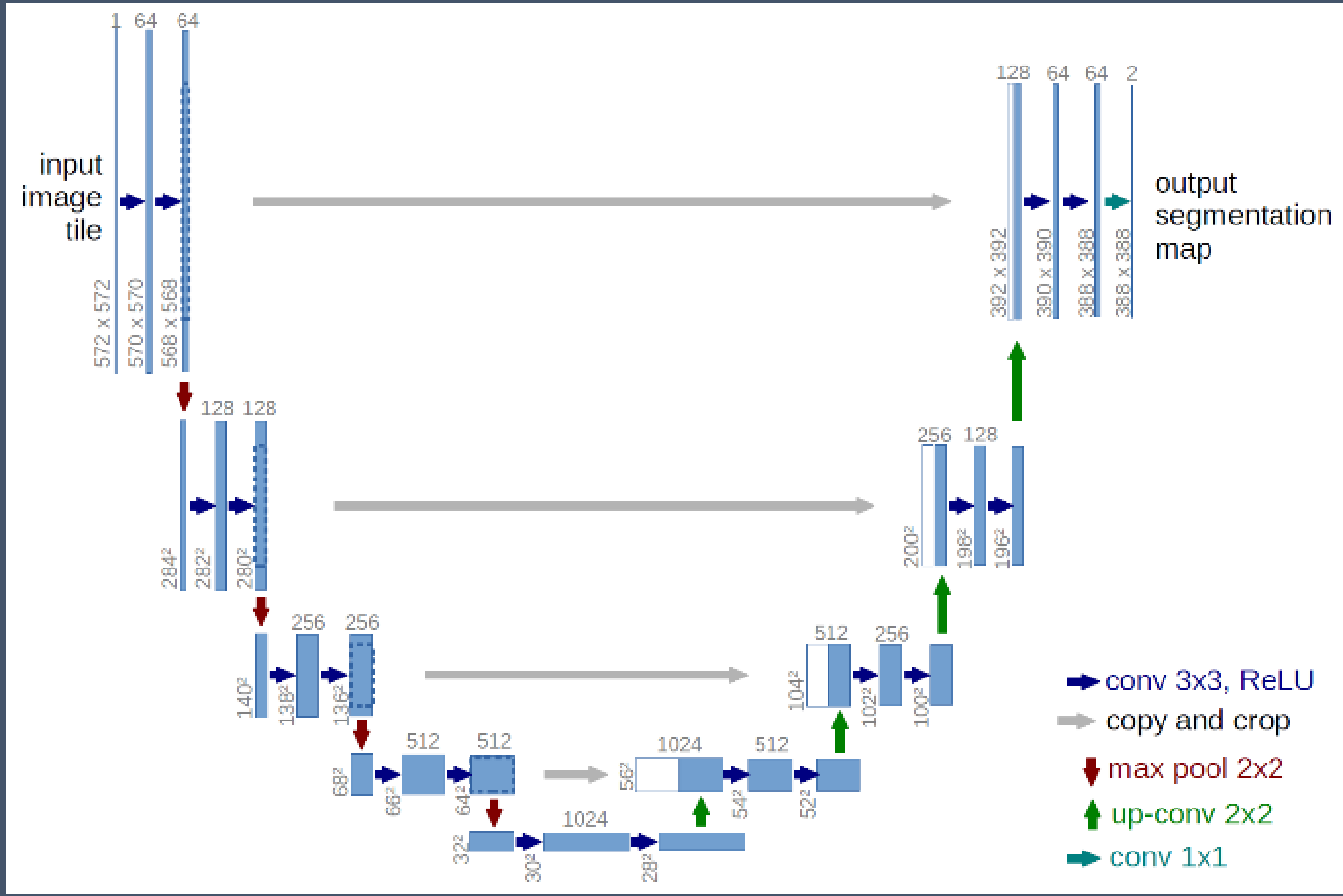
Fully convolutional network
Segmentation

- Add classification layers in upper layers
- Average results
- Can we do better ?



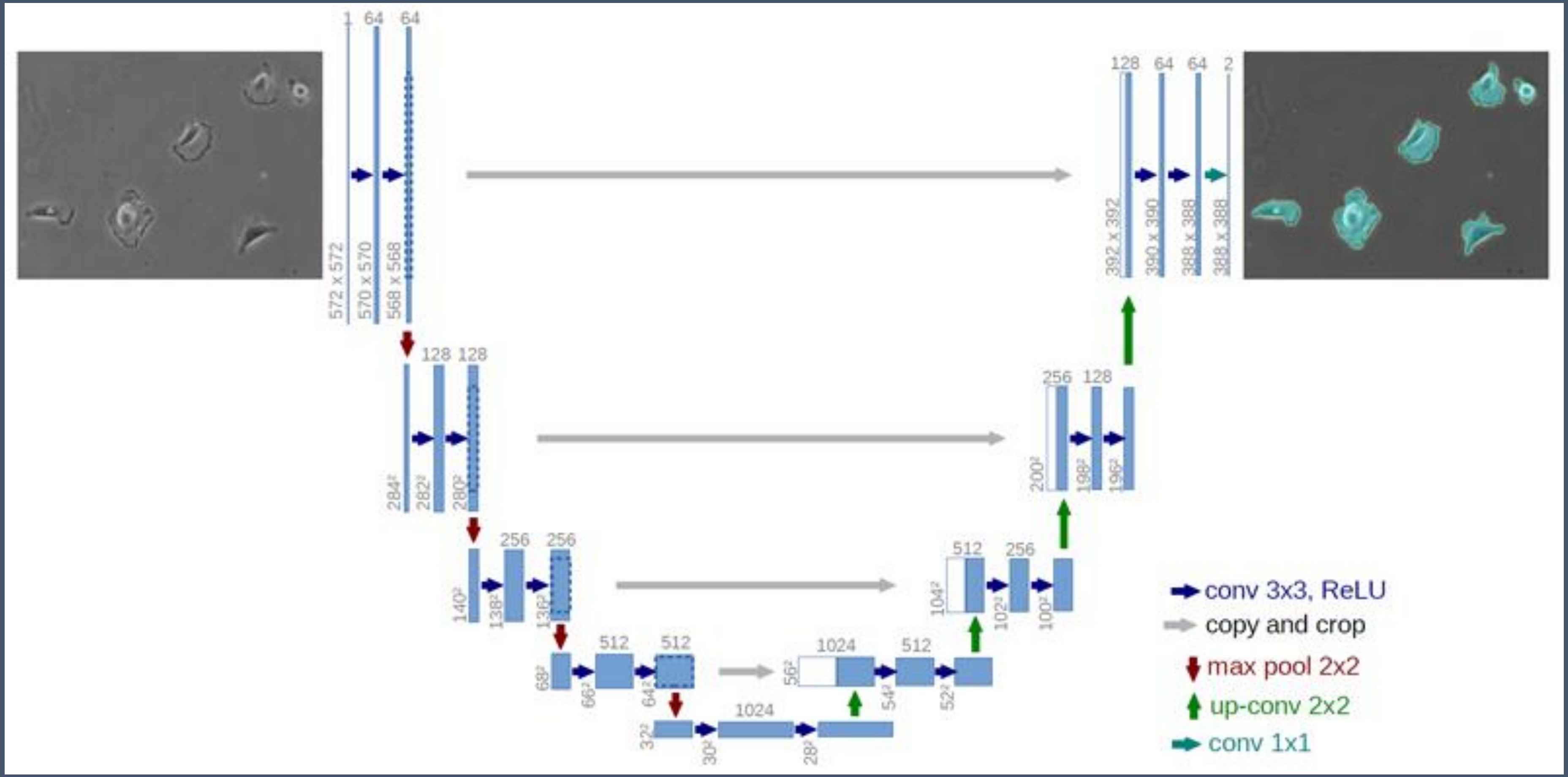
Fully convolutional semantic segmentation
U-Net

- Based on VGG style blocks
- Valid convolutions
- Skip connections at every resolution
- Cascade of refinements
  - Based on coarse decisions
  - Semantically simpler at the finer levels



O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” MICCAI, 2015

U-Net original results



Data problem revisited

- We have more data, but…
  - Data of particular cases often rare
  - Most data not tagged

- Situation worse in medical data
  - Samples of rare malignancies
  - Segmentation of medical images

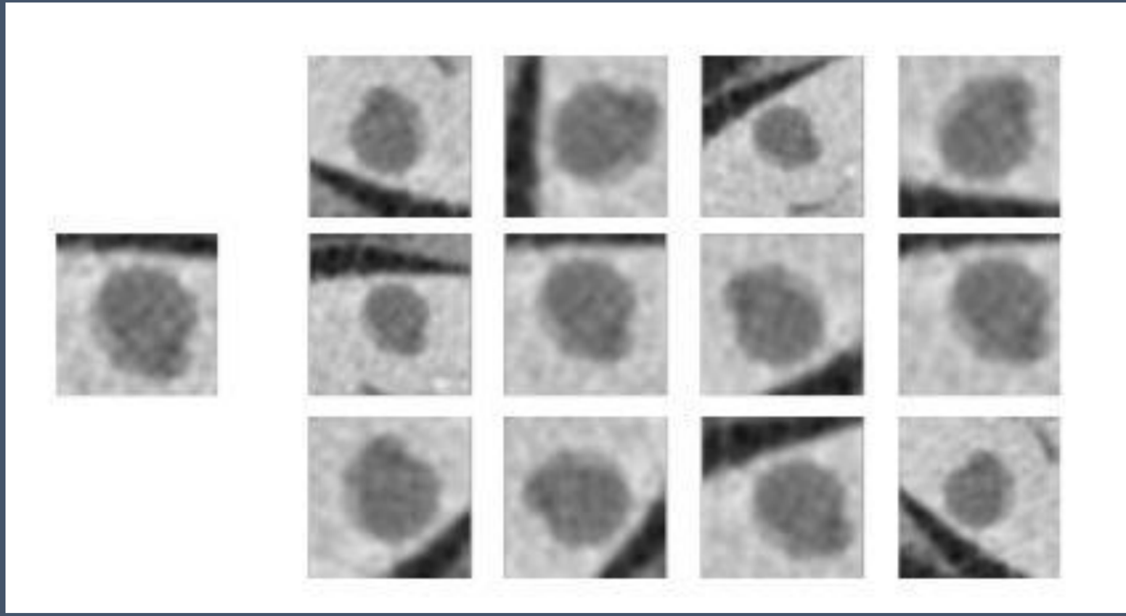
What to do when there is insufficient data

- Data augmentation
- Transfer learning
- Unsupervised learning / Auto-encoders
- GANs

Data augmentation
Classical approach

Combinations of rotation translations flipping scaling and skewing

- Translation less effective in FCN
- Rotation, skew, flip, and scale limited by the relevant / clinical extent
- Effective augmentation bounded (~x10)
- Cannot replace real variability:
  - Age / size / gender / acquisition HW / operator / protocol



GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification
Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan

Data augmentation

Computer graphics

In many cases data is simulated e.g. from models

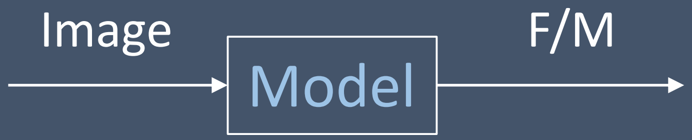
Use real images:
Add lesions from real images to real images of healthy skin

Yunzhu Li, Andre Esteva, Brett Kuprel, Rob Novoa, Justin Ko, Sebastian Thrun
Skin Cancer Detection and Tracking using Data Synthesis and Deep Learning
NIPS Machine Learning for Healthcare Workshop 2016

Transfer Learning in birdland

Male vs. Female
Colibri

Male vs. Female birds



Adapt:

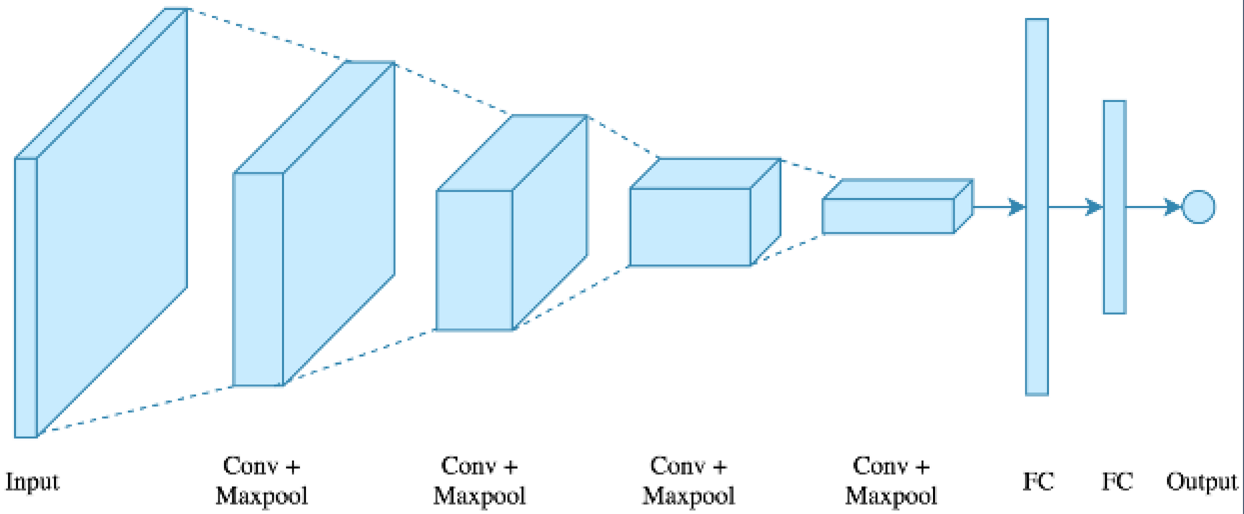
- Keep features (colors)
- Adapt rules (threshold)

What will be the features for mammals

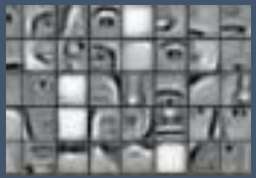
- Size / weight / length of hair

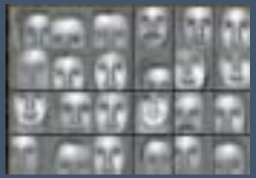
Convolutional layers

Semantics



"Sara"

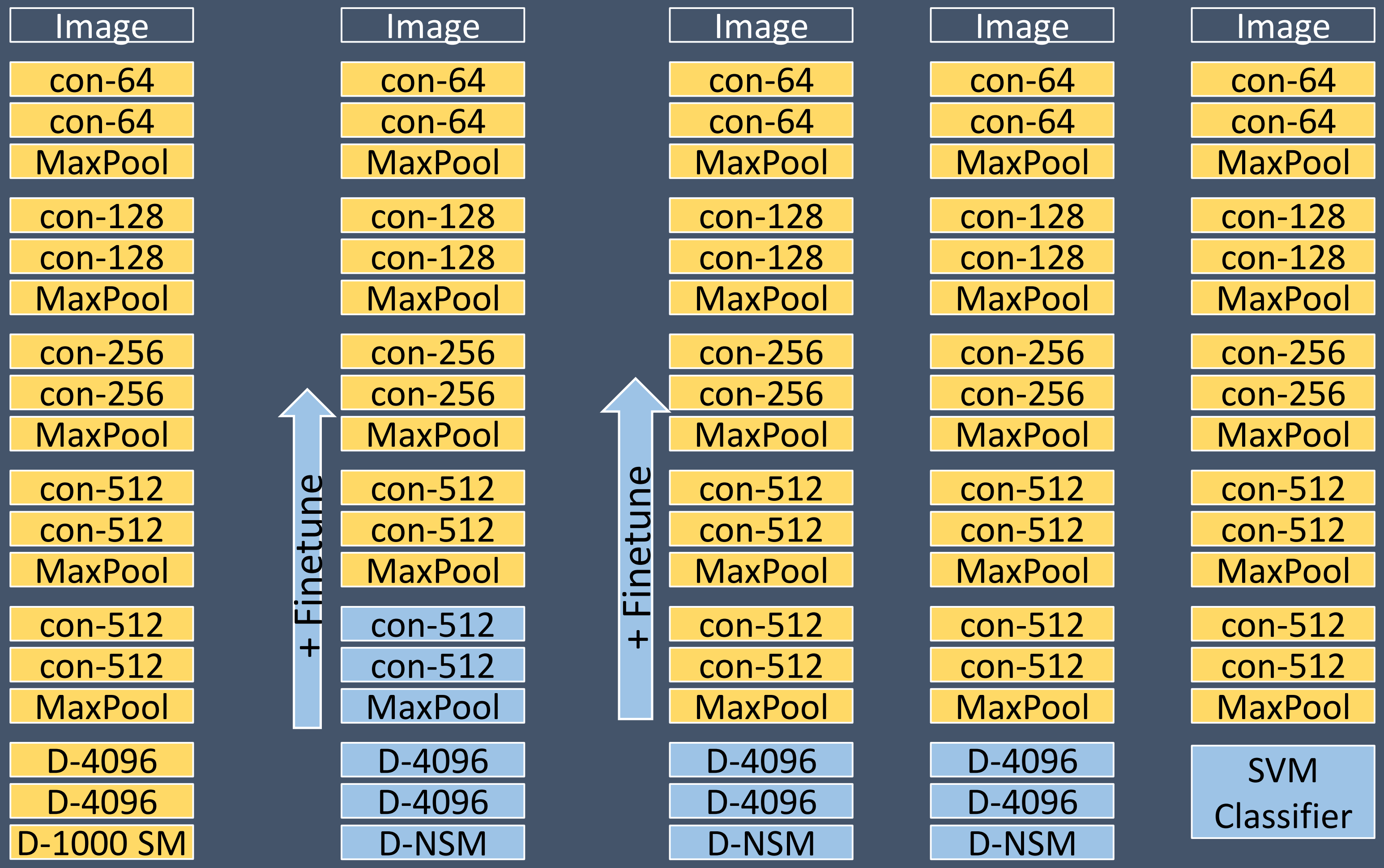




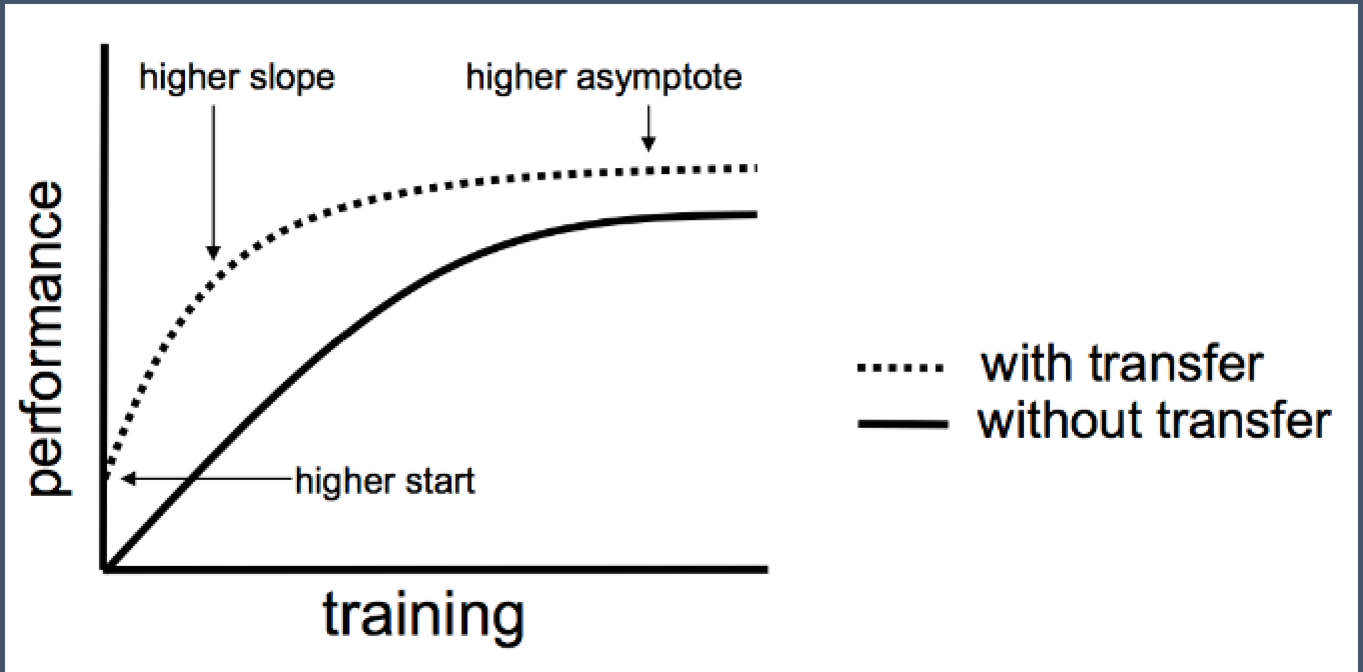
Transfer learning
VGG16 as an example

<u>Medical imaging task</u>

- Very small data:
  - Use top layer as classical features
- Small data
  - Freeze features
  - Train classifier part
- More data
  - As above + finetune features
- Medium data
  - Start from a shallower feature layer
- The different the domain is, transfer becomes less effective



Transfer Learning
Idealized effect of transfer learning



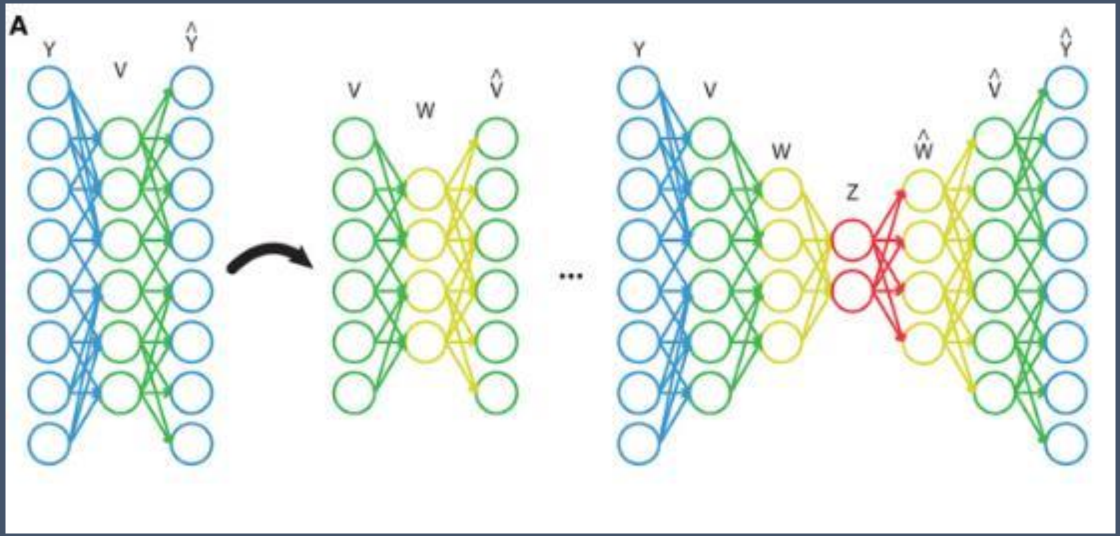
For a real analysis of the effects of transfer learning alternatives for Medical Imaging Analysis:

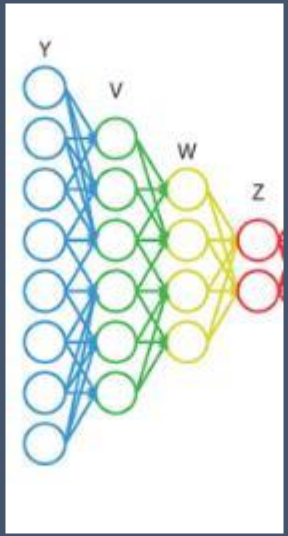
N. Kajbakhsh, J. Y. Shin, S. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. "Convolutional neural networks for medical image analysis: Full training or fine tuning?" IEEE Trans. on Medical Imaging, 2016

Stacked Autoencoders

- Use individual pixels as tags
- Narrow hidden layer to prevent the trivial solution
- Repeat the process (stack)
- You have trained two sub-nets:
  - Encoder
  - Decoder
- Transfer the encoder to your desired application

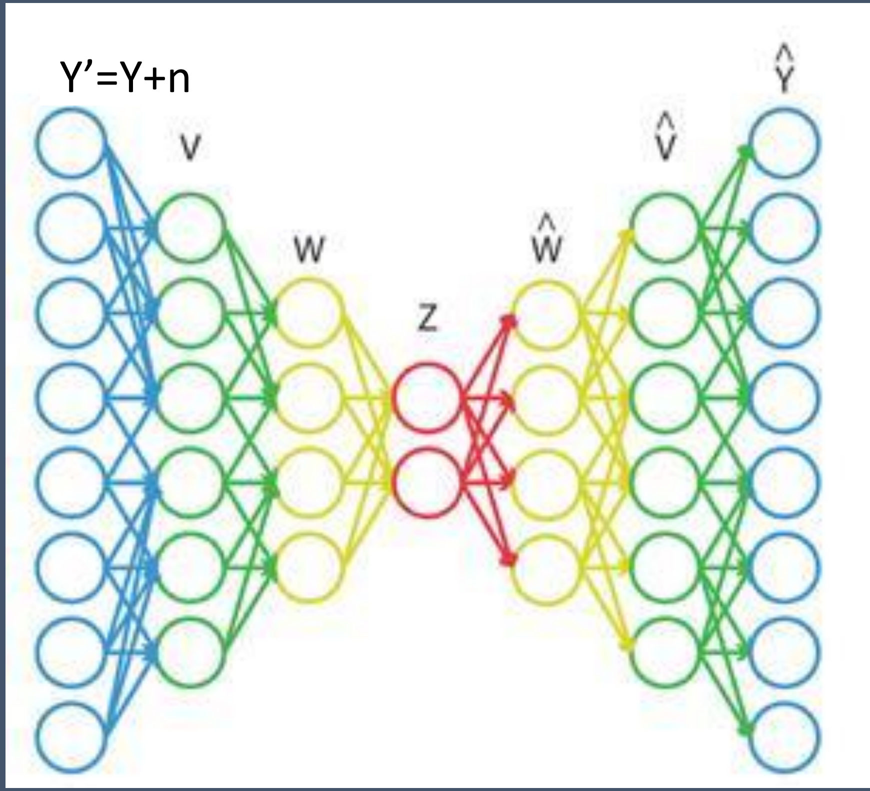
- Trained on your data





Denoising autoencoders

- Estimate clean images from noisy images
- As easy to generate ground truth for as regular autoencoders
- Not as ‘trivial’ a task as regular autoencoders
- Denoising is a valid application in and of itself
- Encourages semantically meaningful features



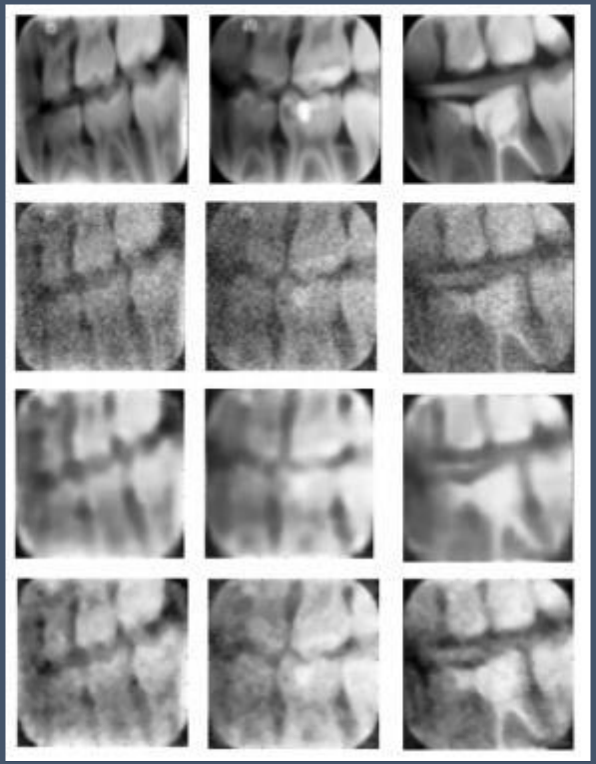
Denoising autoencoders

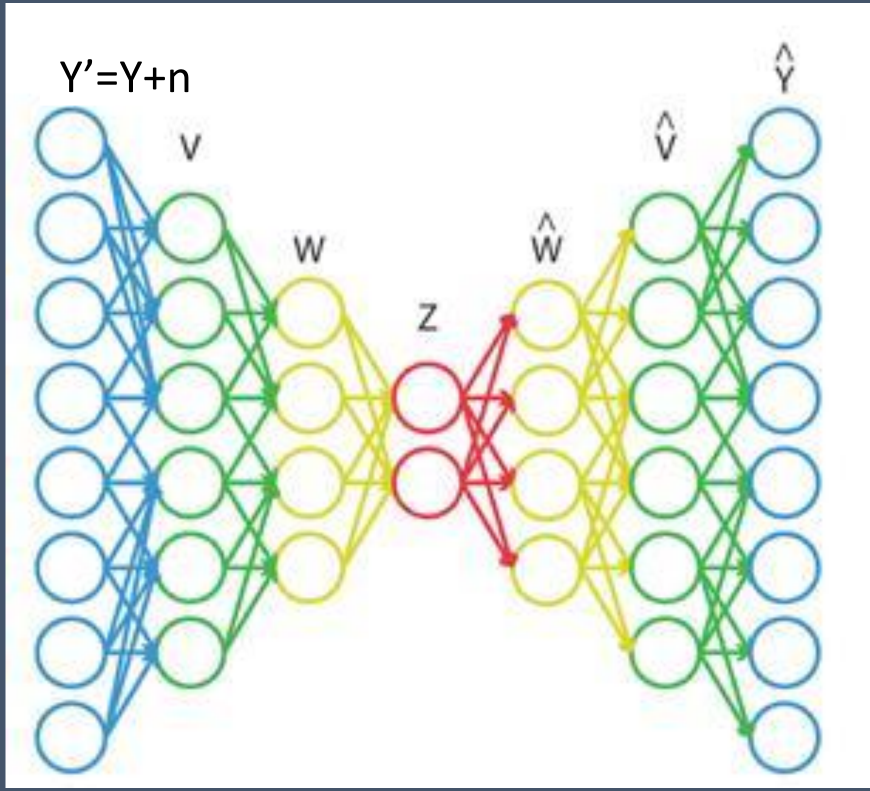
Original image

Noisy version

Denoised via autoencoder

Denoised via Median filter

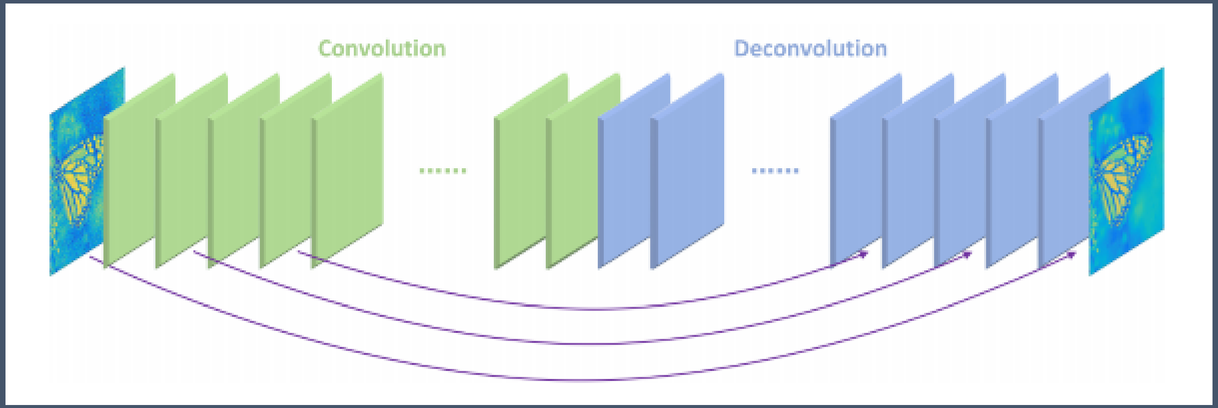


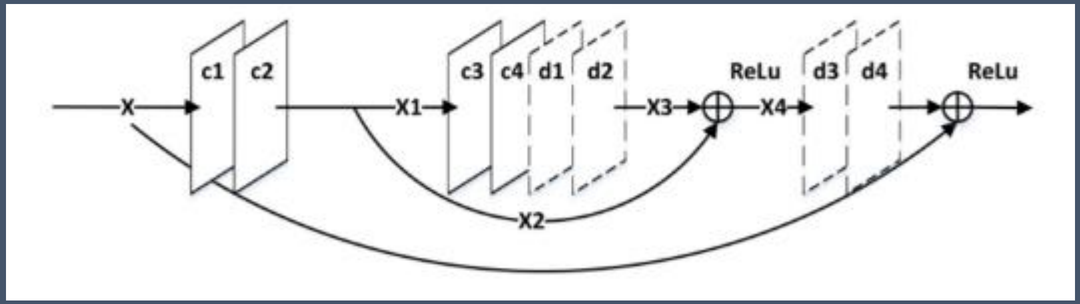


Lovedeep Gondara, Medical image denoising using convolutional denoising autoencoders, Arxiv 2016

Restoration autoencoders

- Estimate clean images from degraded images
  - Noise
  - Down-sampling (super res)
  - JPG artifacts
- Similar to U-Net, but skip layers are summed rather than concatenated
- Wider direct and indirect applicability

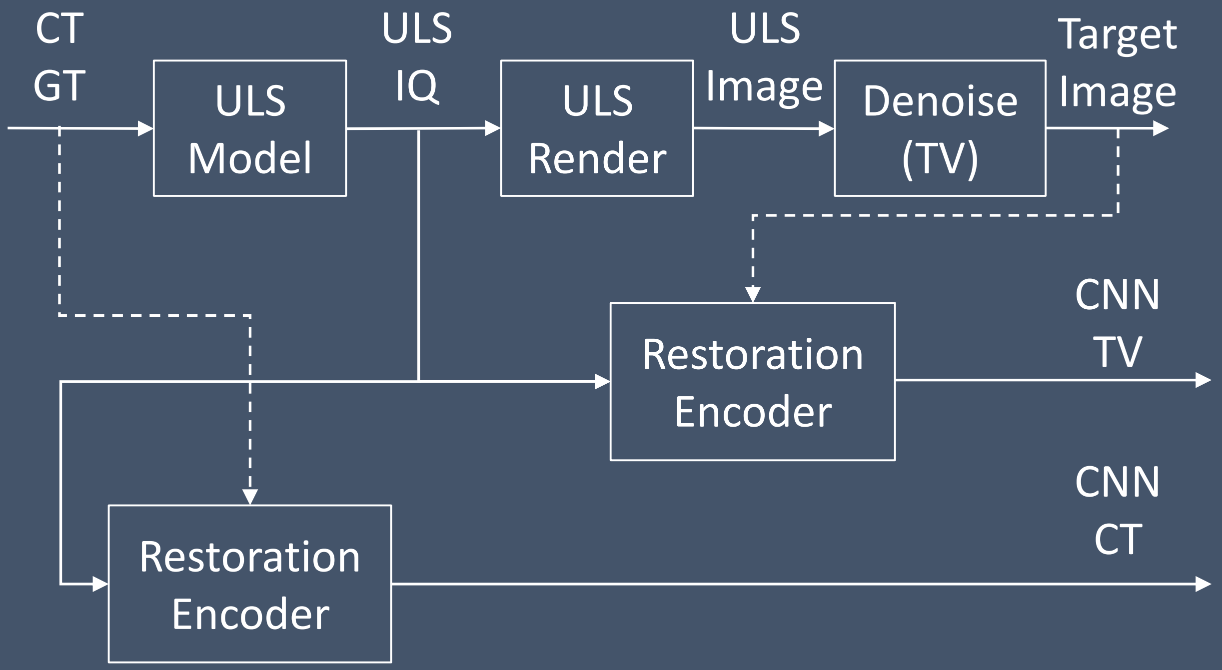


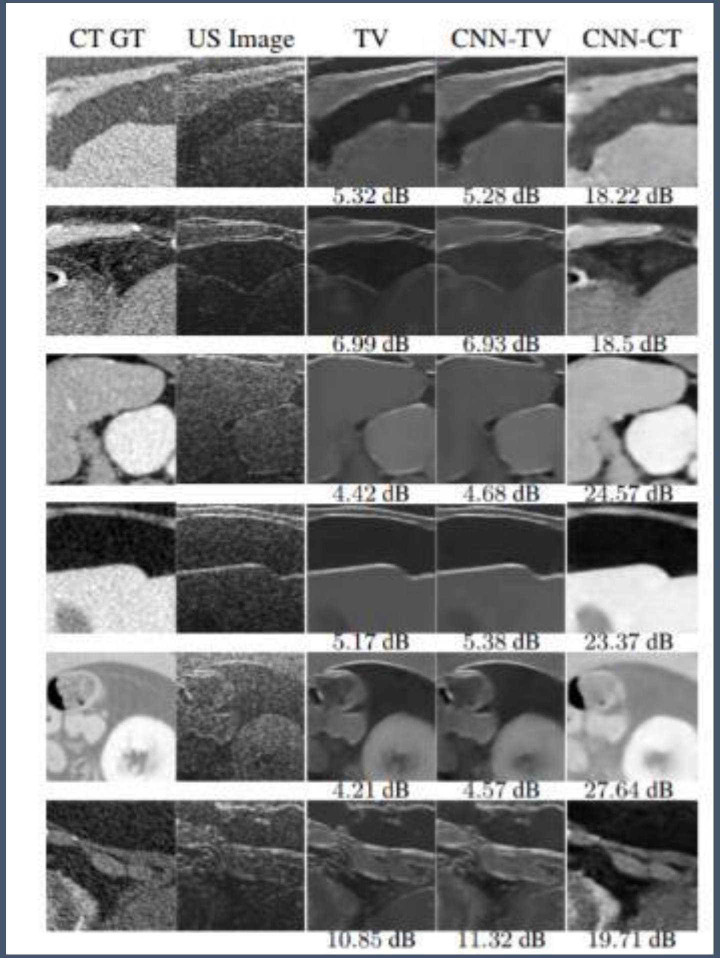


X. J. Mao, C. Shen, and Y. B. Yang,
“Image restoration using convolutional auto-encoders with symmetric skip connections,” CoRR, 2016.

Restoration encoders

Cross modality mapping

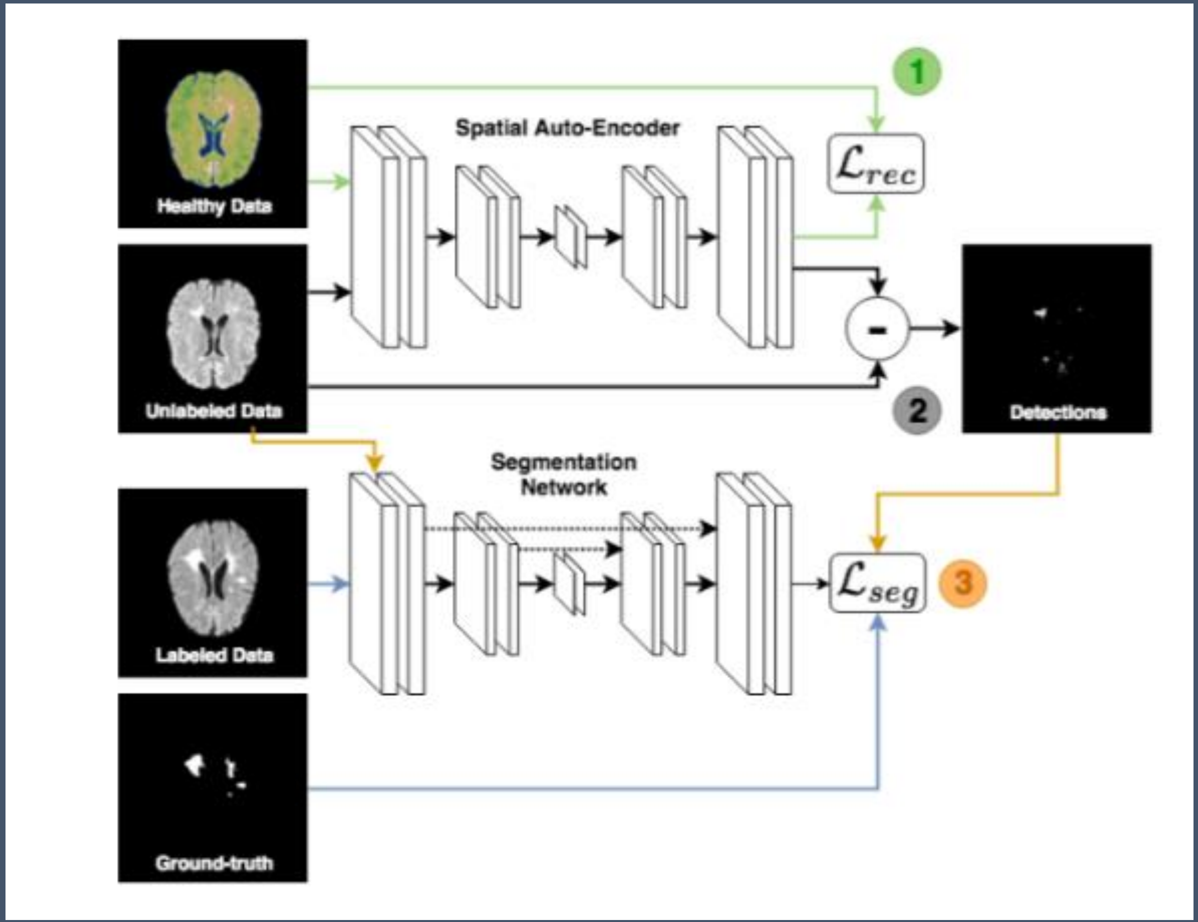




S. Vedula, O. Senouf, A. M. Bronstein, O. V. Michailovich, and M. Zibulevsky, Towards CT-Quality Ultrasound Imaging Using Deep Learning, ArXiv 2017

Semisupervised learning (anomaly detection)

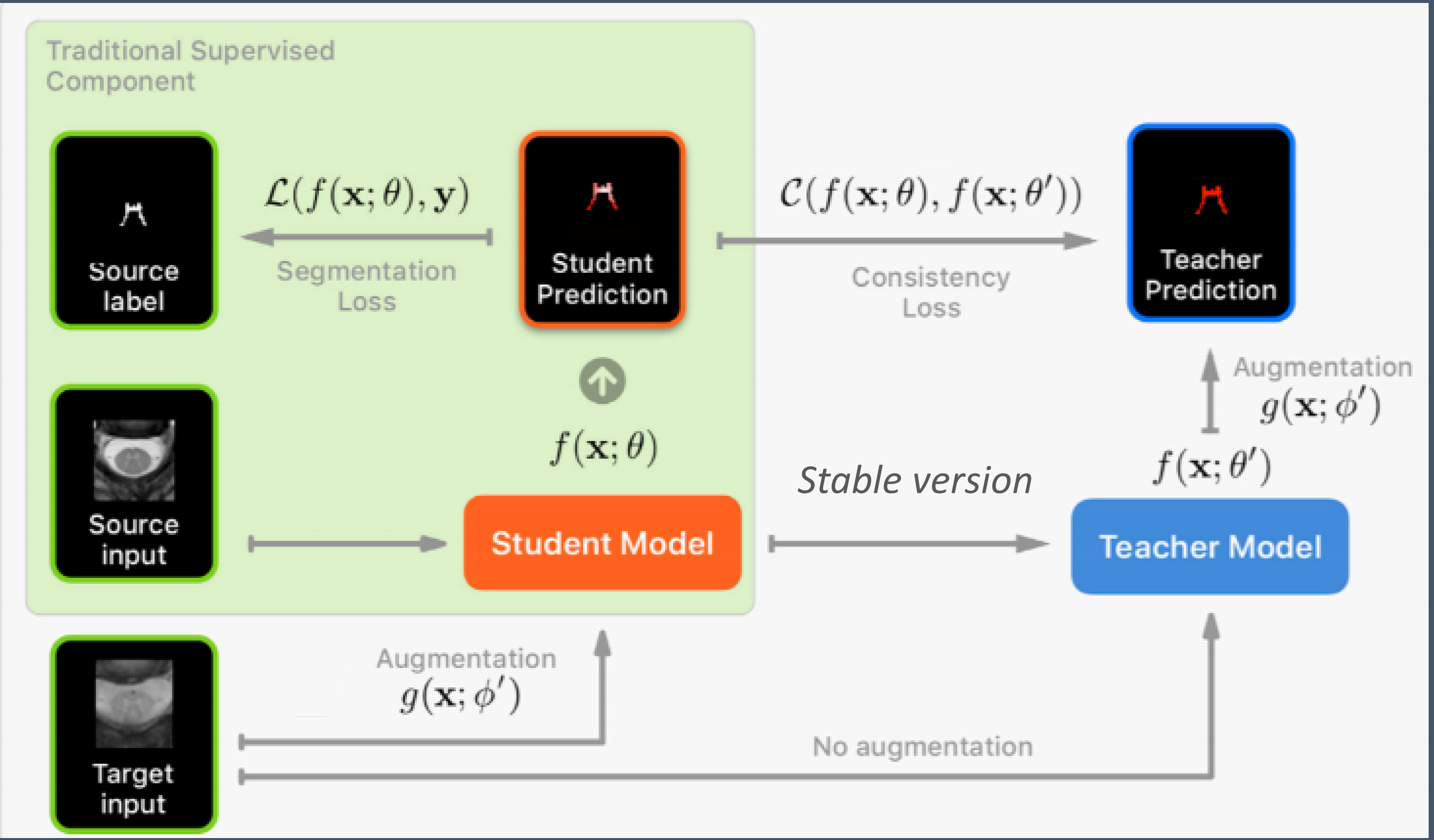
1. Train autoencoder for healthy samples
2. Anomaly detection for unknown samples
3. Train a segmentation NN on few ground truth labels and otherwise anomaly labeles



C. Baur, B. Wiestler, S. Albarqouni, N. Navab,
Fusing Unsupervised and Supervised Deep Learning for White Matter Lesion Segmentation, MIDL 2019

Semisupervised learning (consistency)

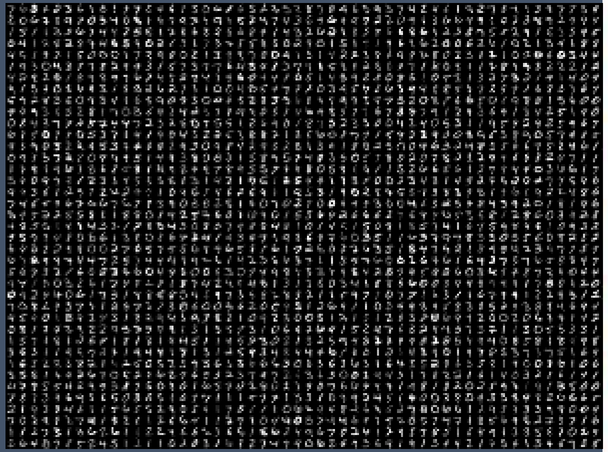
1. Start regular supervised training for ‘Student’
2. Occasionally, update a Stable version as ‘Teacher’
3. Unknown samples will be
   1. Augmented pre-student
   2. Augmented post-teacher
4. Student needs to be consistent with teacher

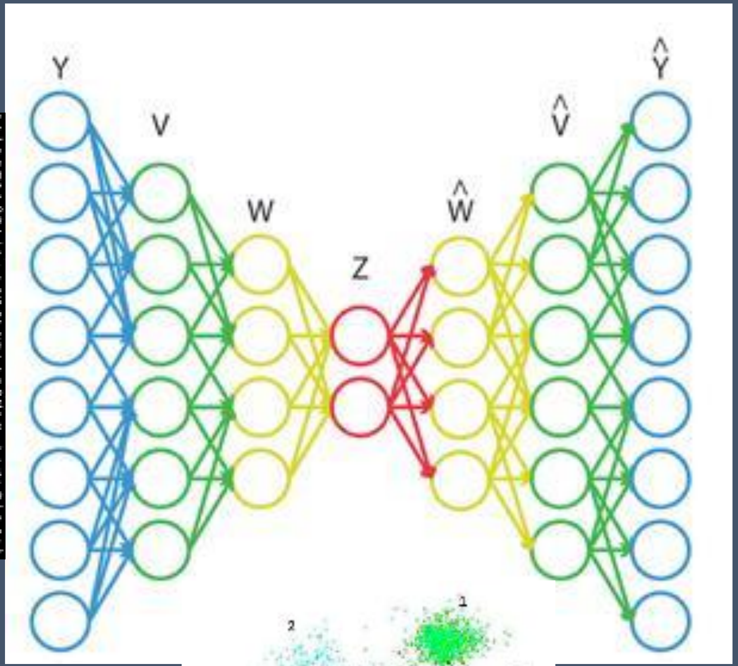


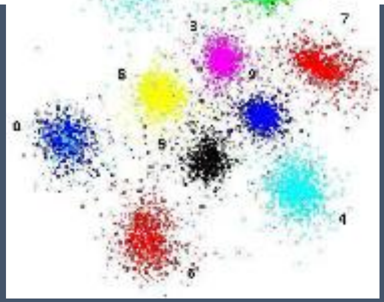
C.S. Perone, P. Ballester, R.C. Barros, J. Cohen-Adad,
Unsupervised domain adaptation for medical imaging segmentation with self-ensembling, NeuroImage 2019

Revisiting encoders

- Imagine we trained an autoencoder for MNIST images
  - Encoding into 2 features
  - All digits map into 2D
- We can use the encoder as a pre-trained feature extractor for MNIST tasks
- We can use the decoder as a generator for unseen handwritten digits
- Are there alternative ways to train a generator?



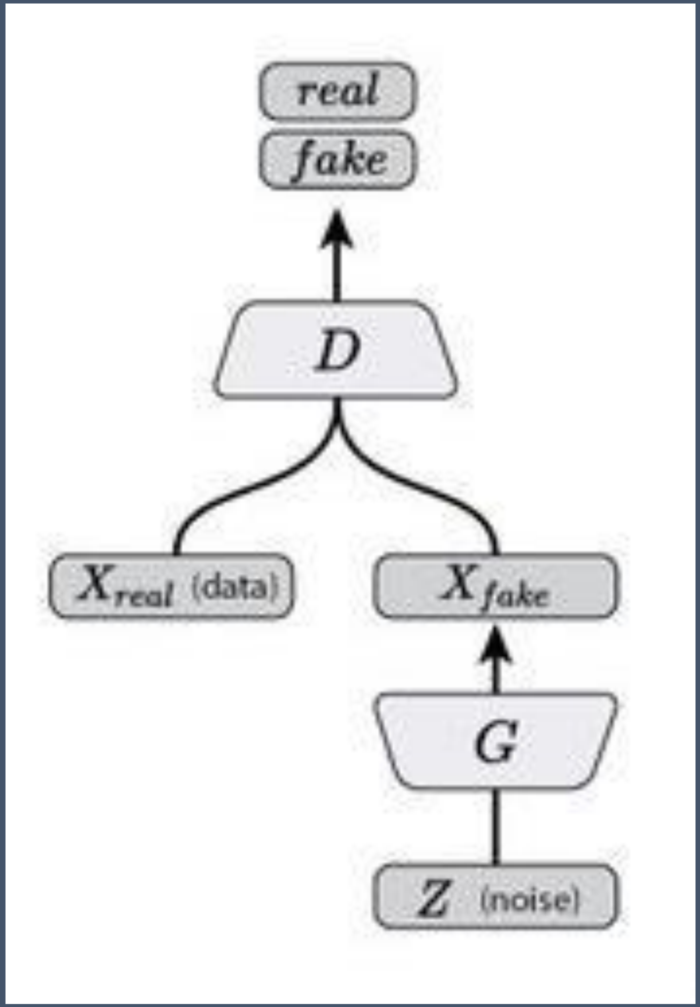




Generator: Most inputs map to a valid output.
Valid output: Indistinguishable from true images

GAN – Generative Adversarial Networks

- It should be really easy to train a discriminator to identify true images from fake images
- Will need a small *representative* set of true images
- Use the discriminator to train a generator
- Train both consecutively to improve both



I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, Generative adversarial nets. NIPS, 2014.

GAN – Generative Adversarial Networks

Objective function

$$\min_{W_g} \max_{W_d} \left[E_{x\sim X}\{\log D_{W_d}(x)\} + E_{z\sim Z}\left\{\log\left(1 - D_{W_d}\left(G_{W_g}(z)\right)\right)\right\}\right]$$

Log likelihood that: real is identified

Log likelihood that: fake is identified (1-D)

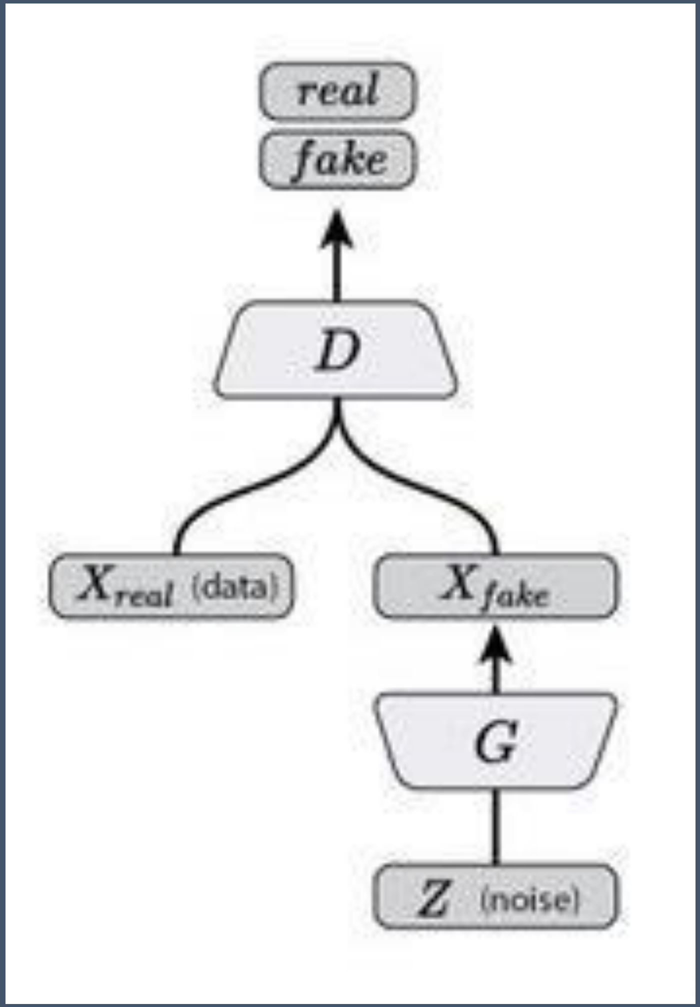
For a batch of m real samples x^i and m noise samples z^i :

Update the Discriminator:

$$W_d \mathrel{+}= \delta \cdot \frac{\partial}{\partial W_d} \sum_{i=1}^{m} \log D_{W_d}(x^i) + \log\left(1 - D_{W_d}\left(G_{W_g}(z^i)\right)\right)$$

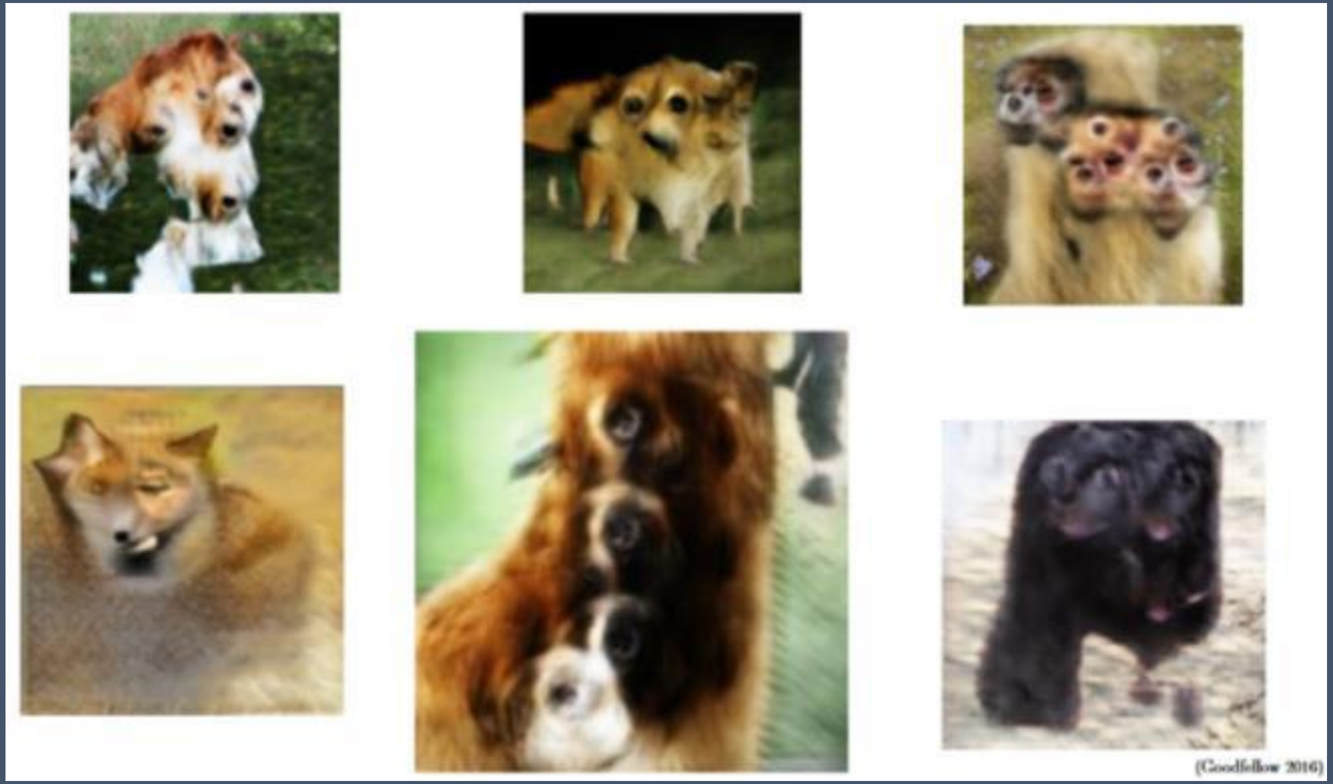
Update the Generator:

$$W_g \mathrel{-}= \delta \cdot \frac{\partial}{\partial W_d} \sum_{i=1}^{m} \log\left(1 - D_{W_d}\left(G_{W_g}(z^i)\right)\right)$$



Problems with GANs

- GAN’s should converge to a Nash equilibrium
- They often stop converging (mode collapses)
- No way to identify convergence
- Even when they converge they are … not perfect
- But…

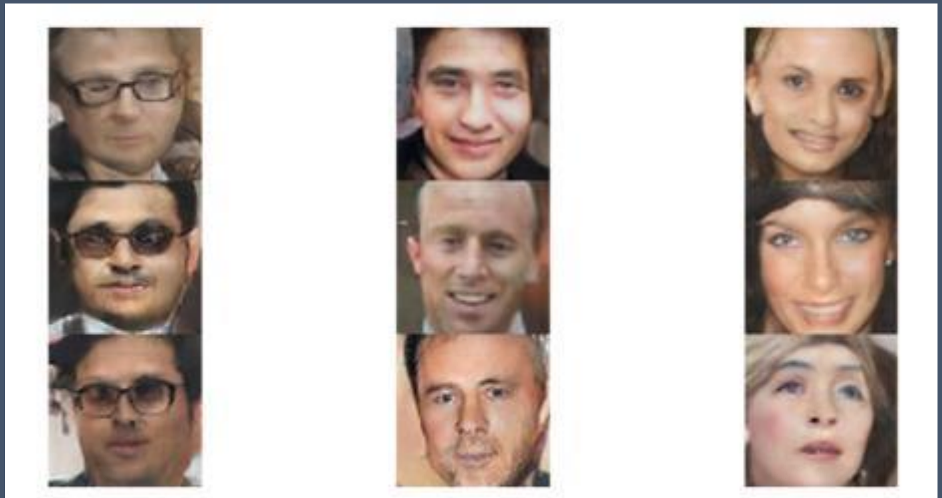


Hallmarks of good embeddings

Smooth interpolation

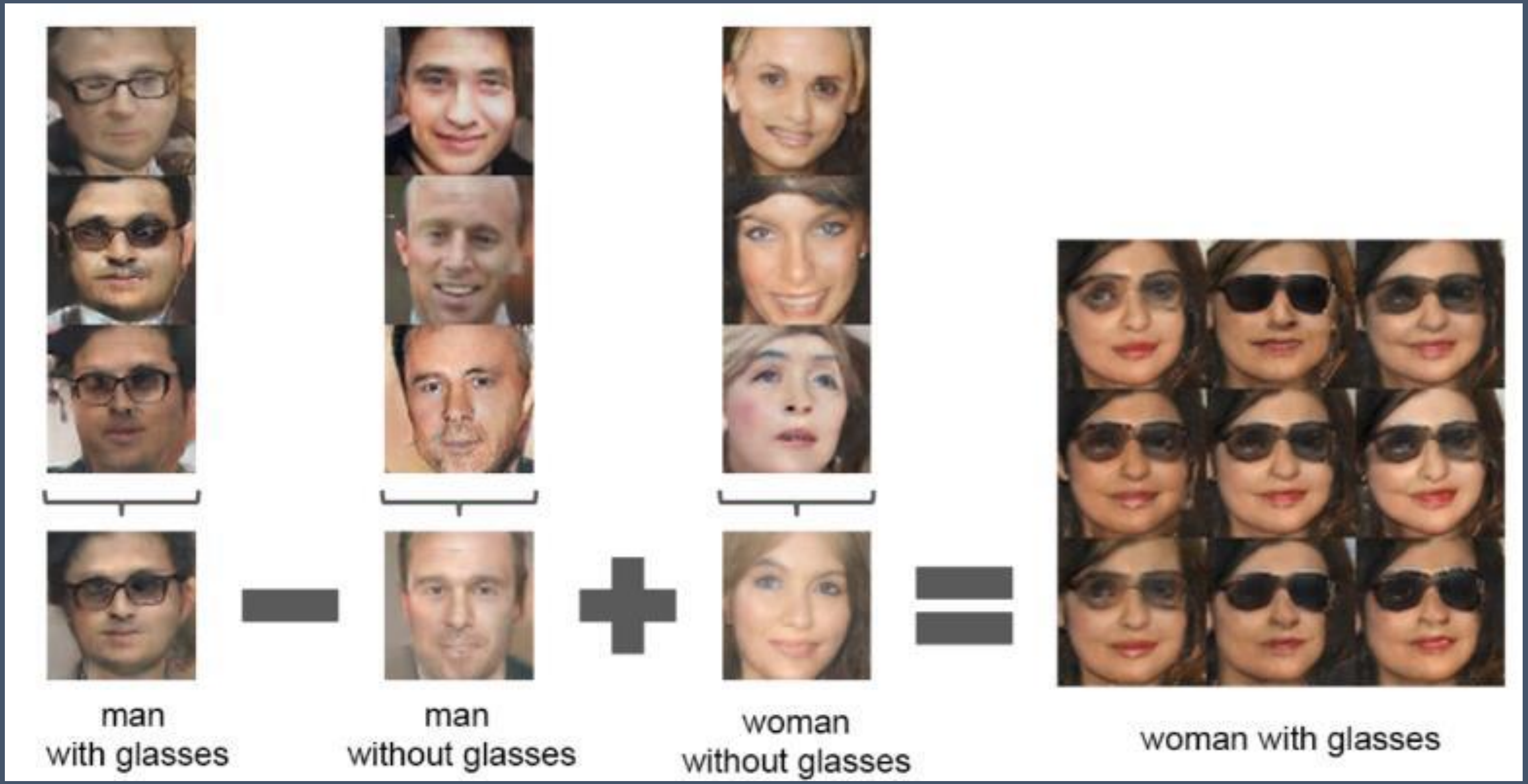
Hallmarks of good embeddings

Semantic arithmetic



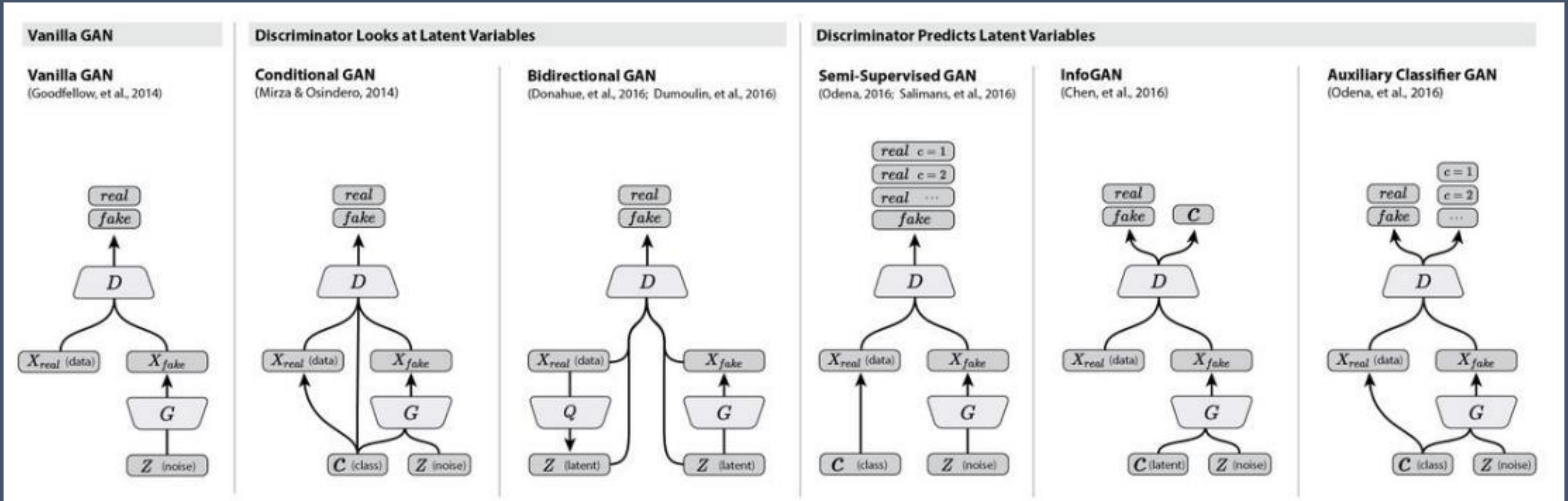
Hallmarks of good embeddings

Semantic arithmetic

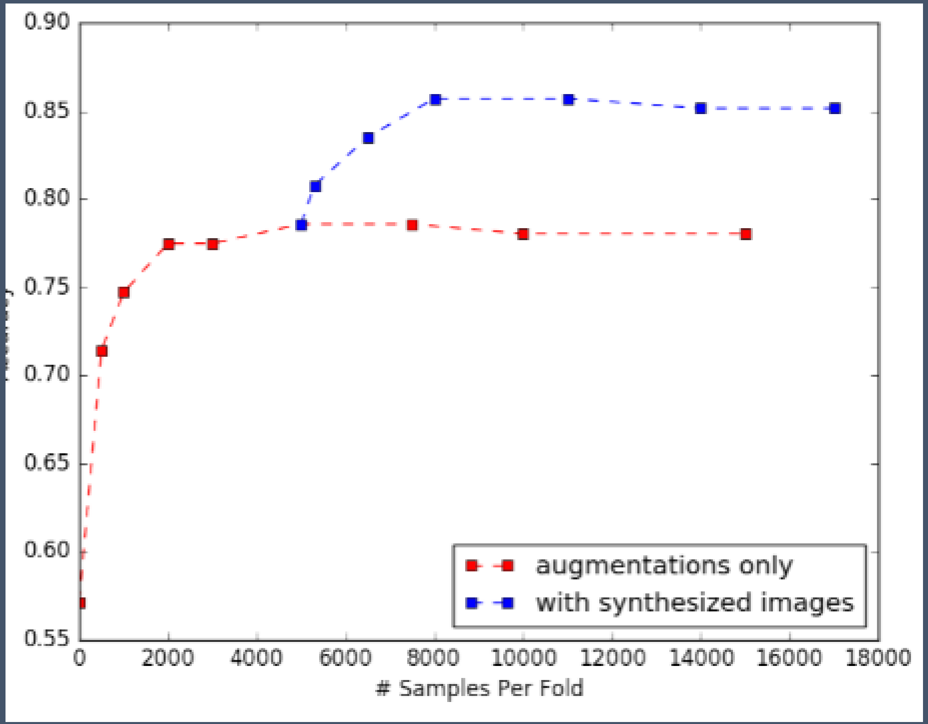


Wave of new GAN architectures

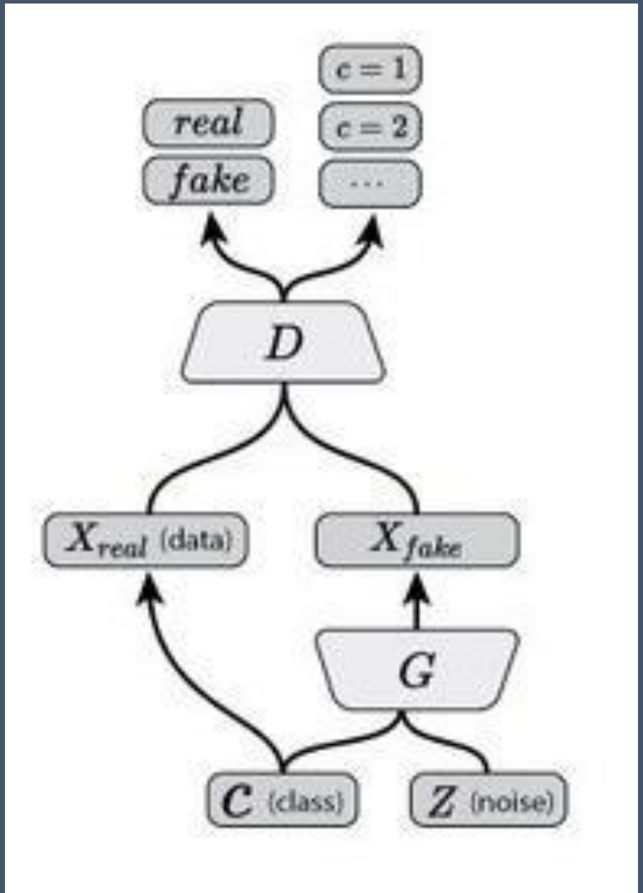
Designed for stability



Auxiliary Classifier GAN
Lesion Classification



M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan,
GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification, 2018



A. Odena, C. Olah, J. Shlens, “Conditional Image Synthesis with Auxiliary Classifier GANs”, 2016

GAN supported Encoder
Cross modality synthesis

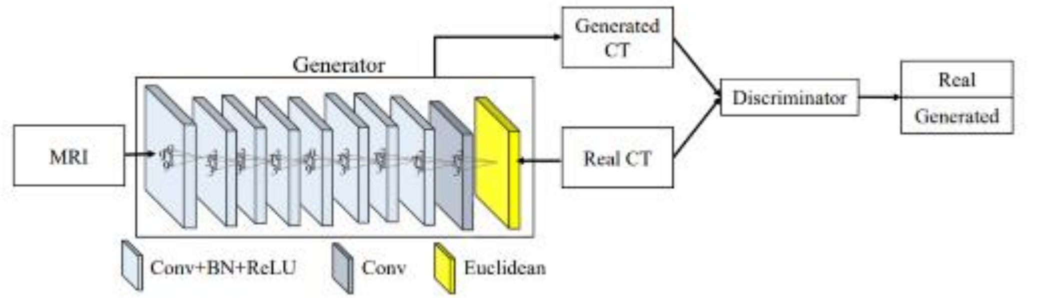
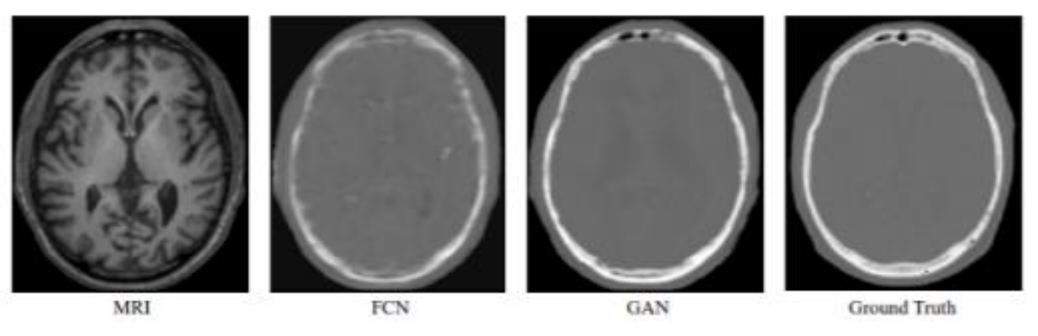
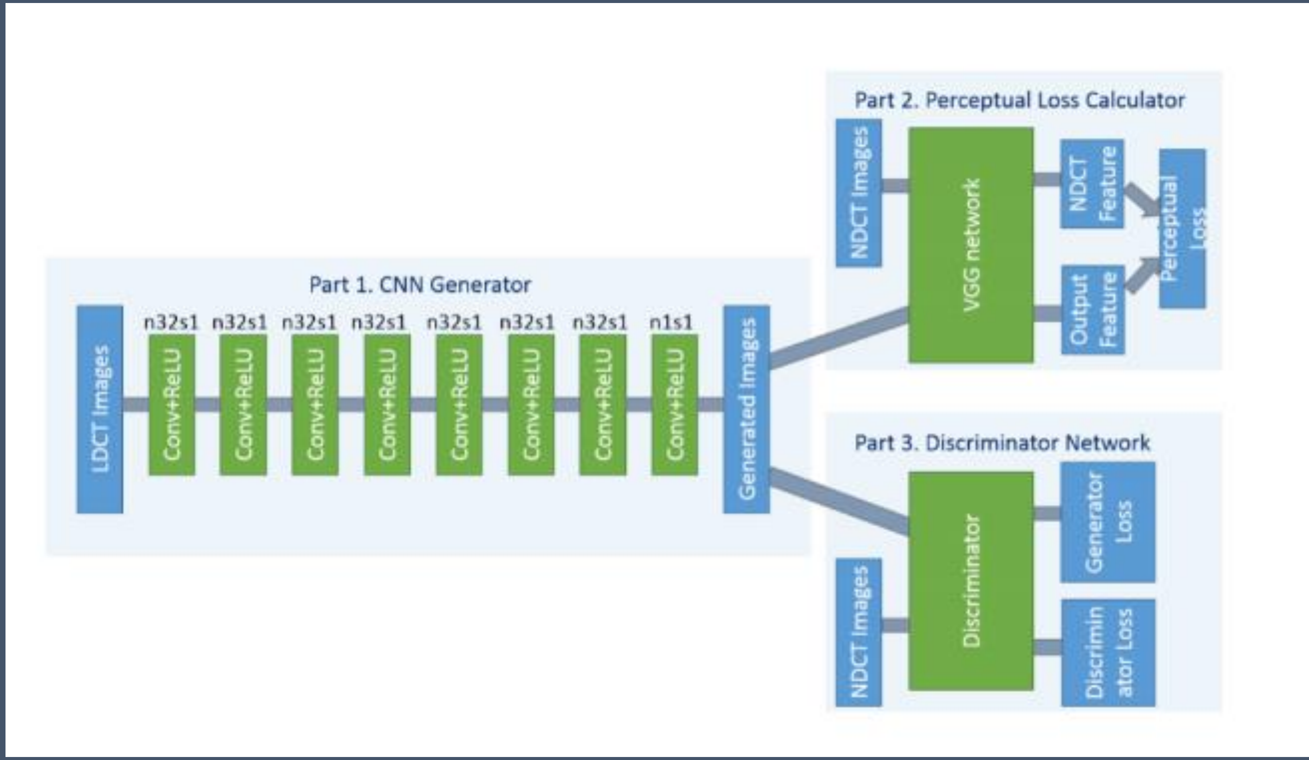


Fig. 2. Architecture used in the Generative Adversarial setting used for estimation of synthetic images.



D. Nie, R. Trullo, C. Petitjean, S. Ruan, and D. Shen, Medical Image Synthesis with Context-Aware Generative Adversarial Networks, MICCAI 2017

GAN supported Perceptual Loss Encoder
Low dose CT reconstruction



Low Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss, Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, Arxiv 2018