01.04.2020

Statistical Methods in AI (CSE/ECE 471)

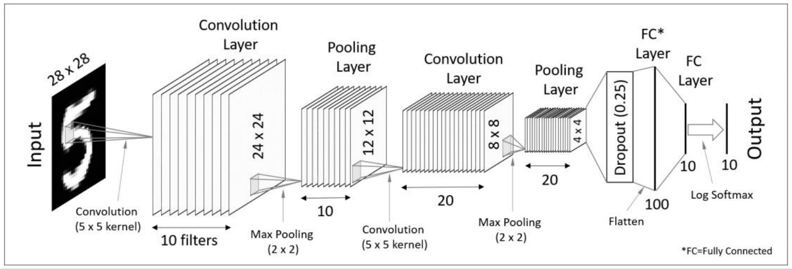
Representation Learning (Siamese Network, Autoencoders)

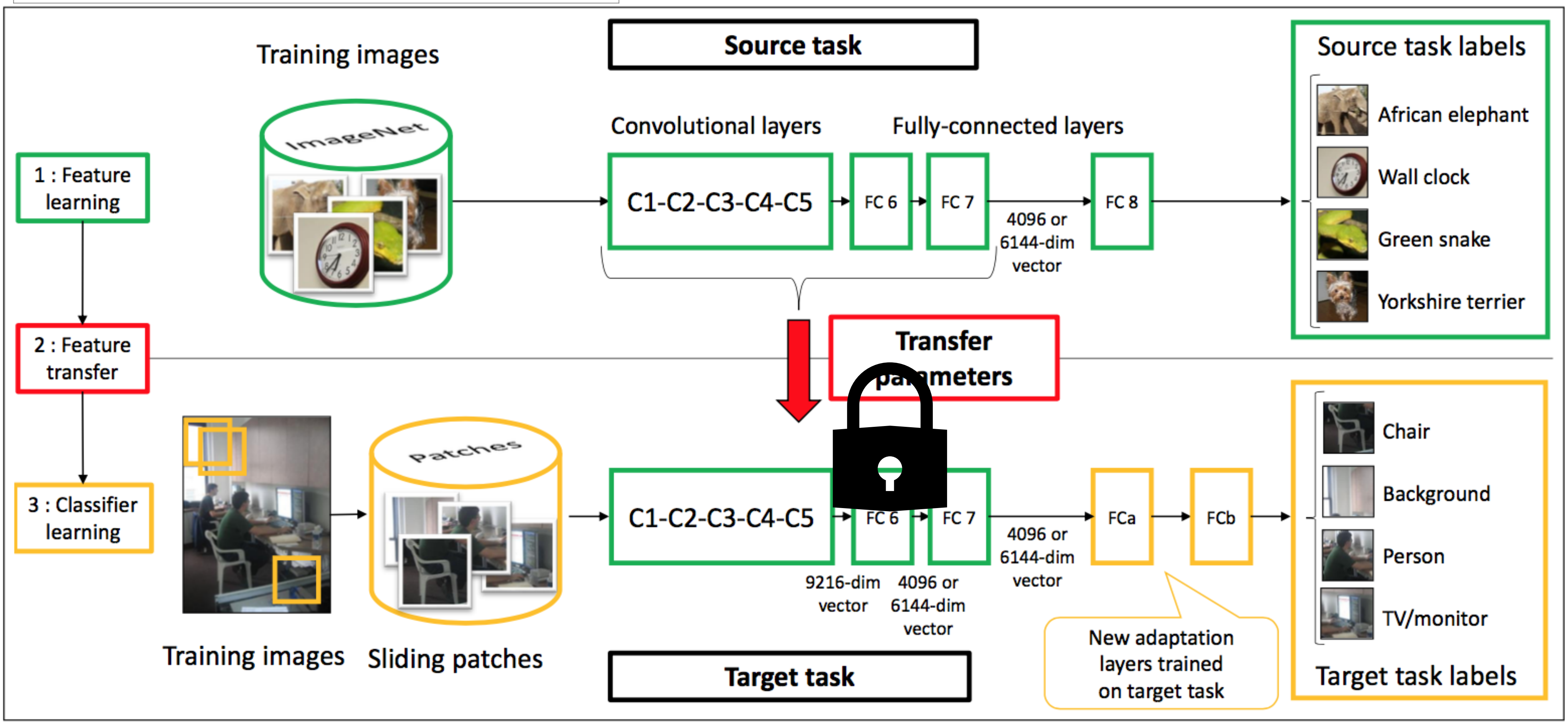
Ravi Kiran (ravi.kiran@iiit.ac.in)

Vineet Gandhi (v.gandhi@iiit.ac.in)

Center for Visual Information Technology (CVIT)
IIIT Hyderabad

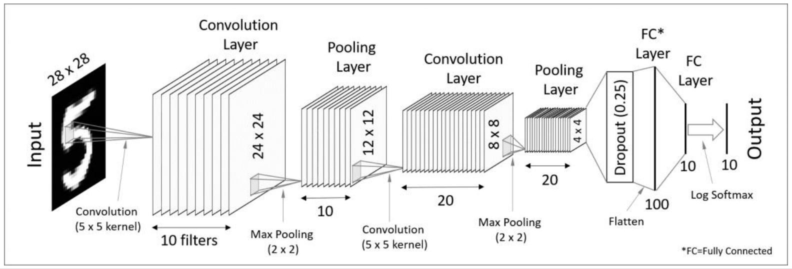
Transfer Learning : Approach-1

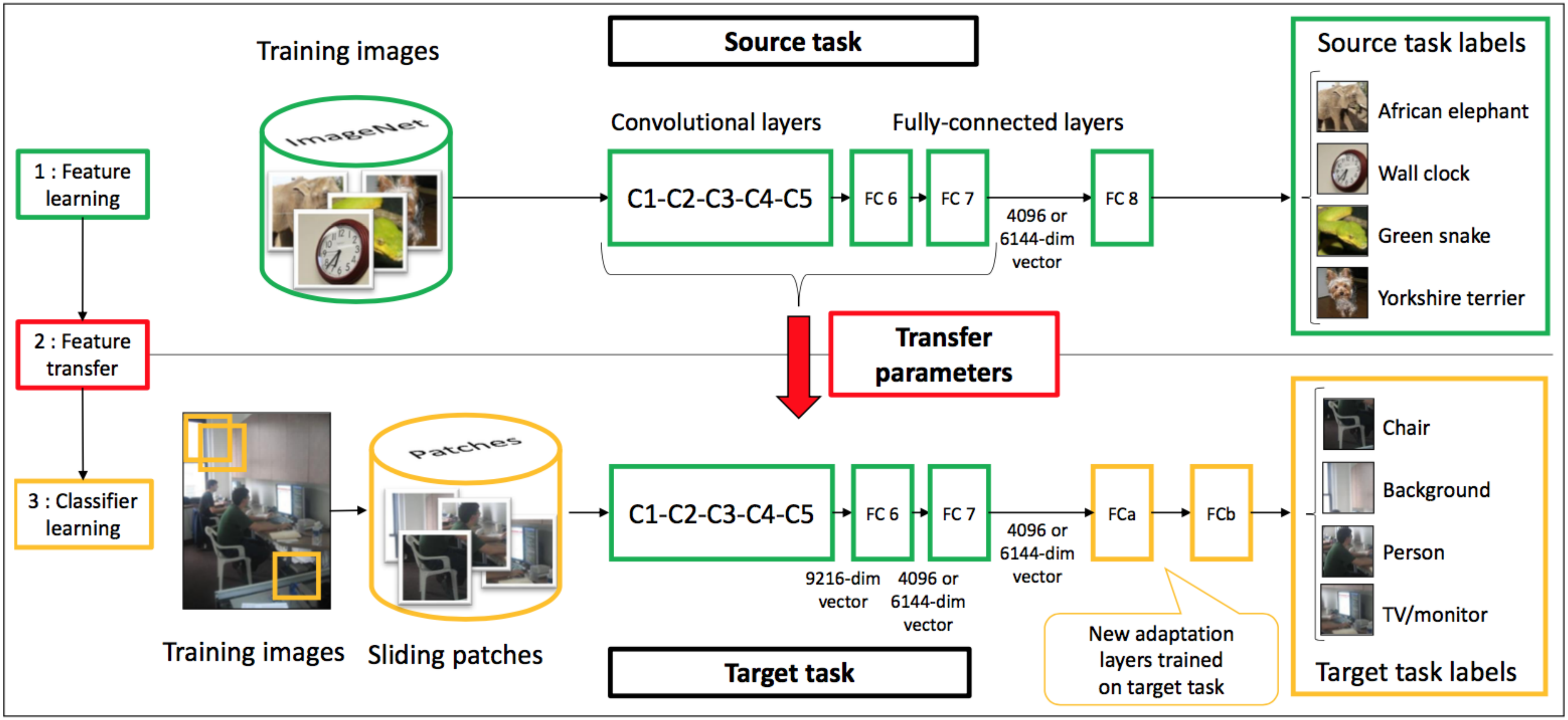




- Learn only weights for newly added layers.
- Ideal when ‘new domain’ data is small in quantity

Transfer Learning : Approach-2

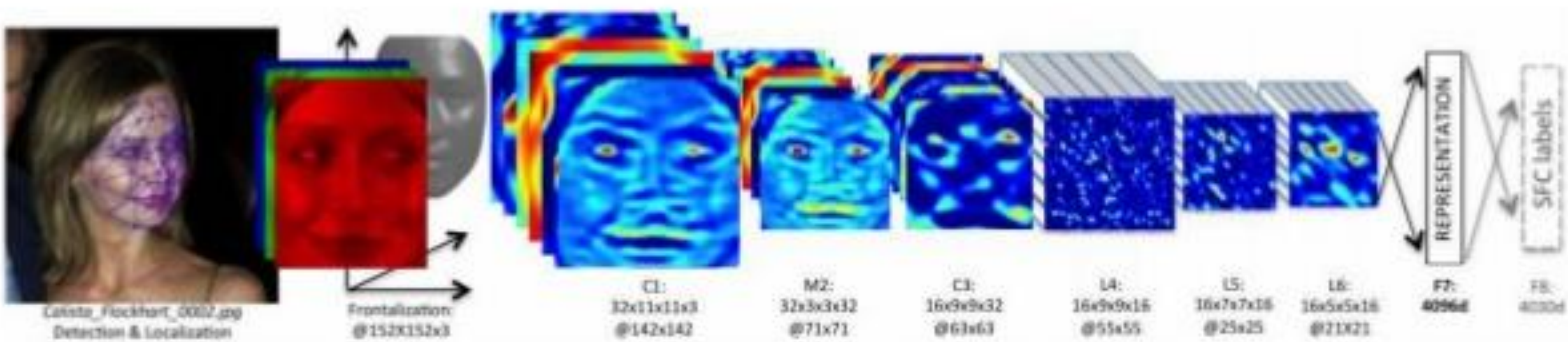




- LR for new layer weights = 10 * source_lr (for bias, 20 * source_lr)
- Ideal when 'new domain' data is reasonably large or domain shift is significant

Classification

Face Identification/Recognition (1:N matching)

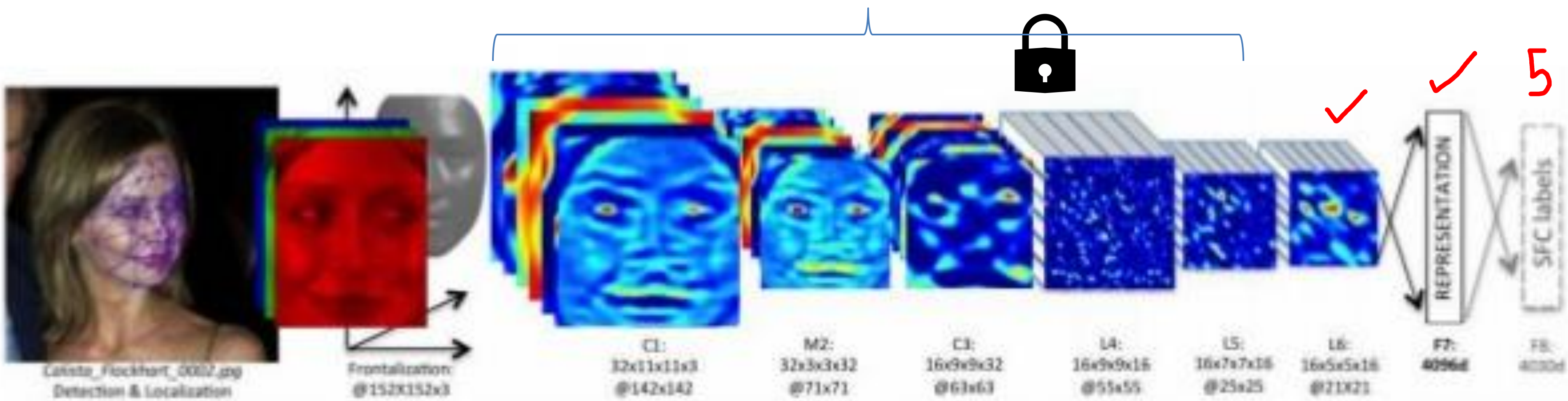


How to reuse DeepFace (trained on celebrities) for another face dataset ?

Classification

Classification

Face Identification/Recognition (1:N matching)



How to reuse DeepFace (trained on celebrities) for another face dataset ?

Ans: Fine-tuning

No-finetuning Classification

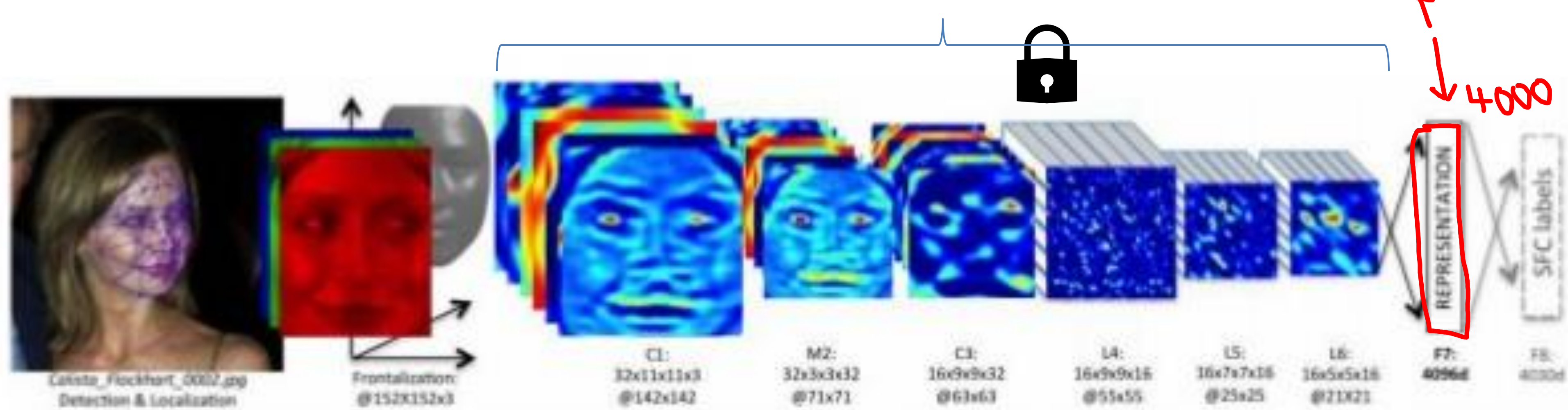
Classification

Face Identification/Recognition (1:N matching)

$f_1^1 \quad f_1^2$

$f_2^1 \quad f_2^2$

$\vdots$



How to reuse DeepFace (trained on celebrities) for another face dataset (without any training) ?

Ans: Use CNN as feature extractor. k-NN on feature representations

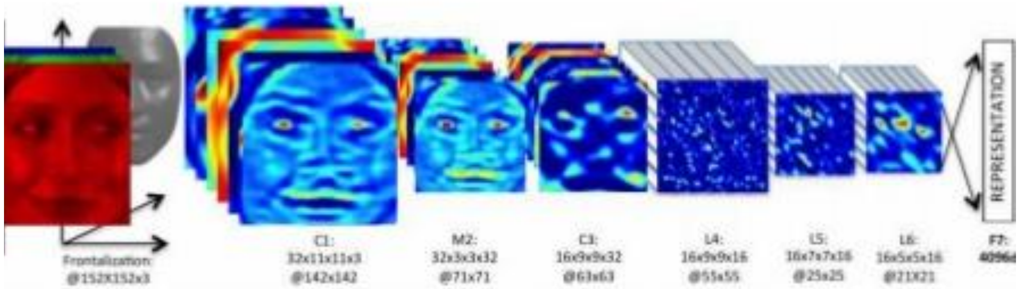
Verification

Face Authentication/Verification (1:1 matching)

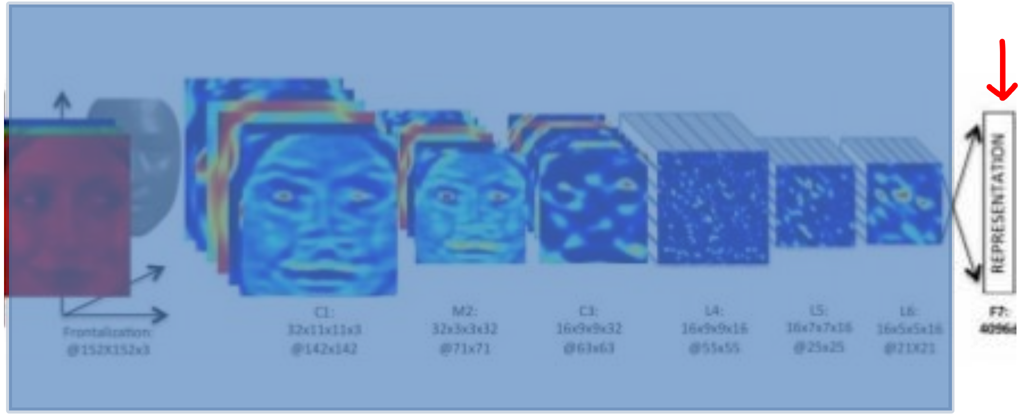
q

d

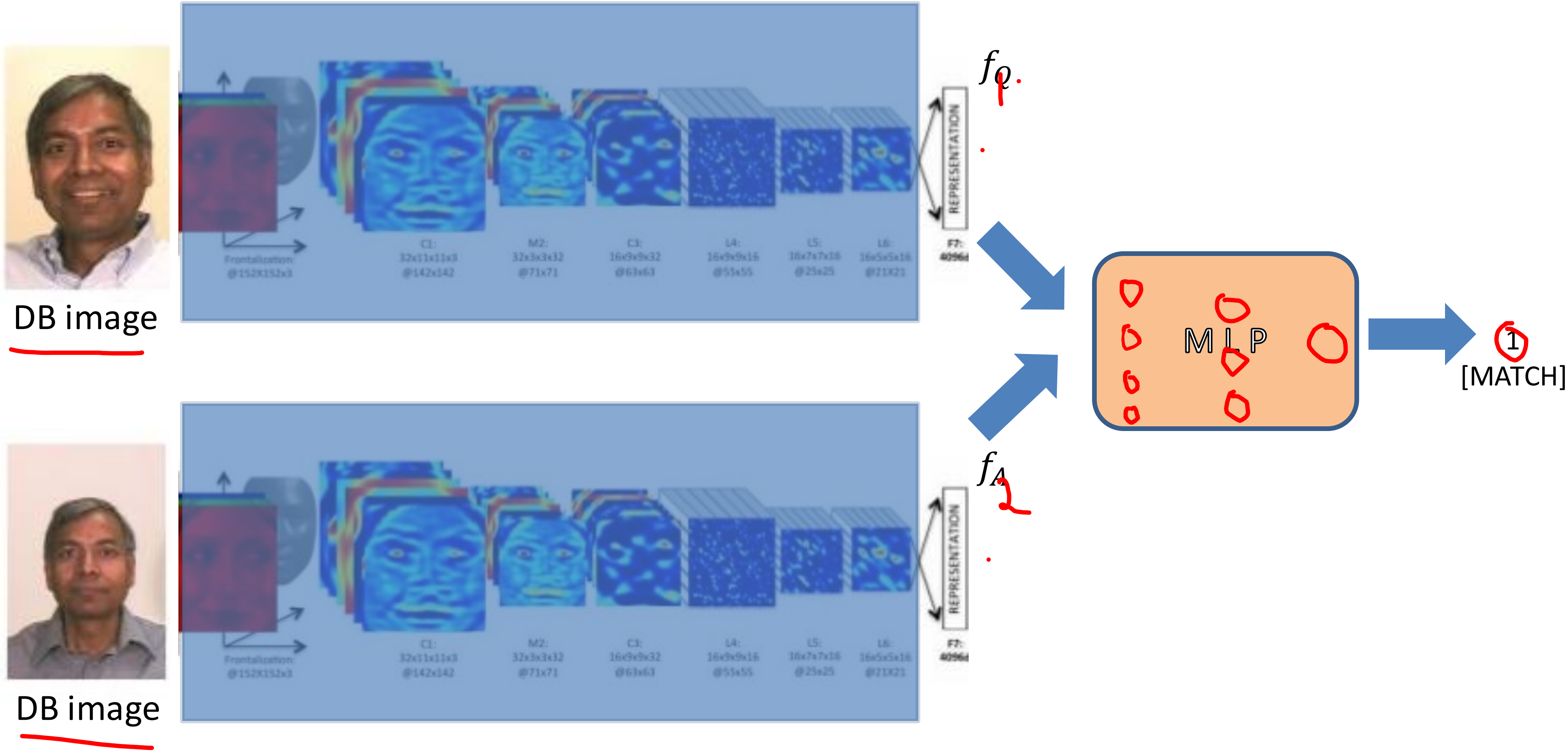
Feature Extraction



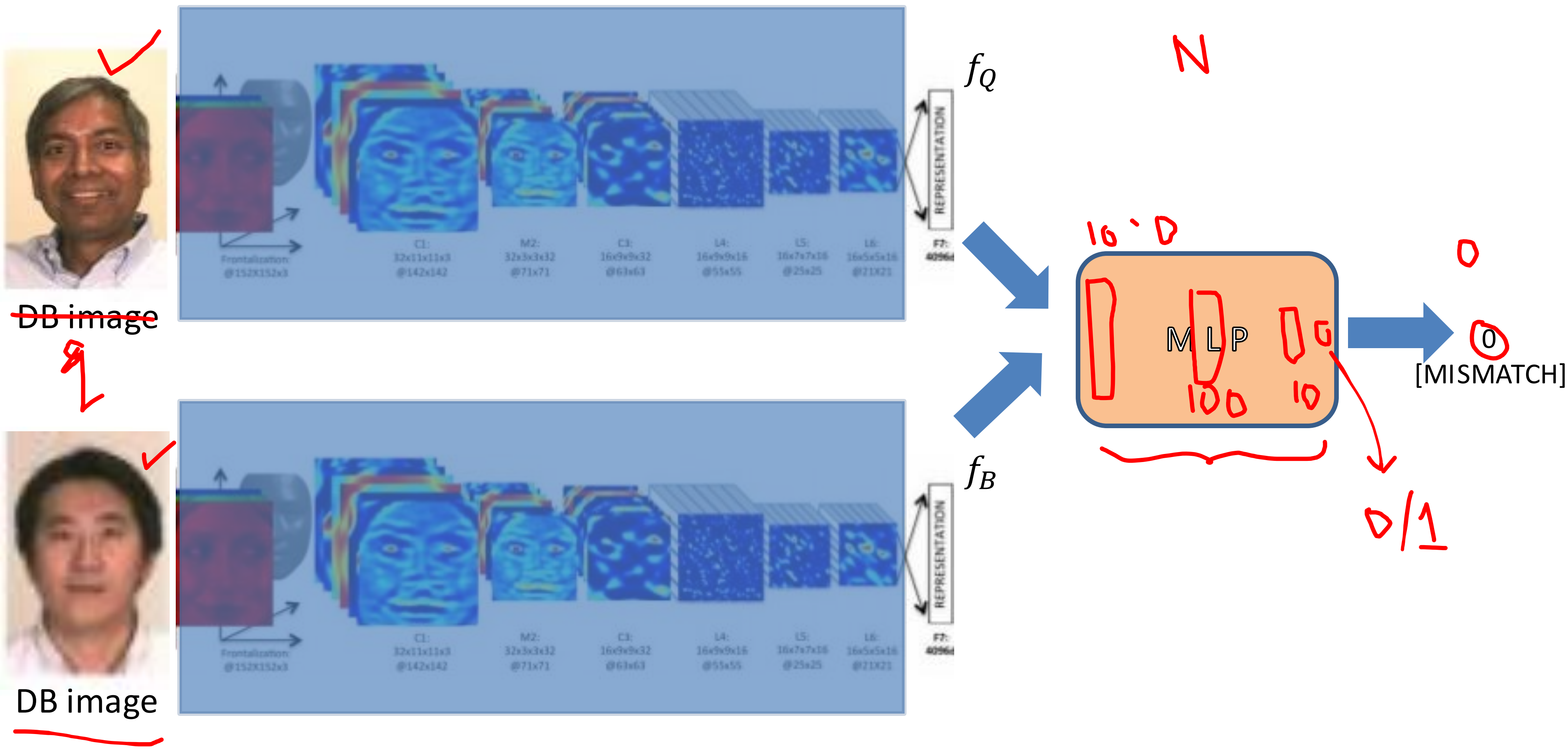
Feature Extraction



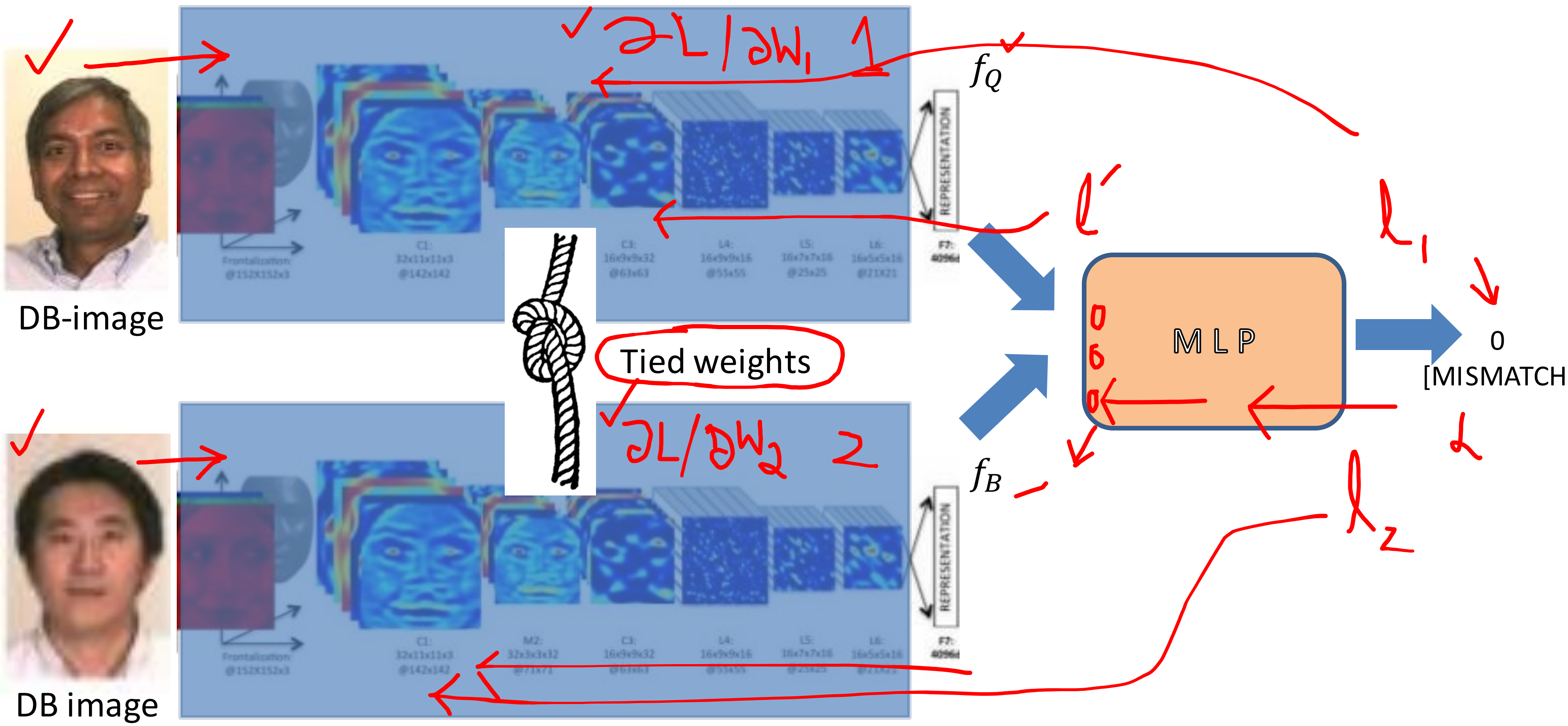
Verification: Approach - 1



Verification: Approach - 1

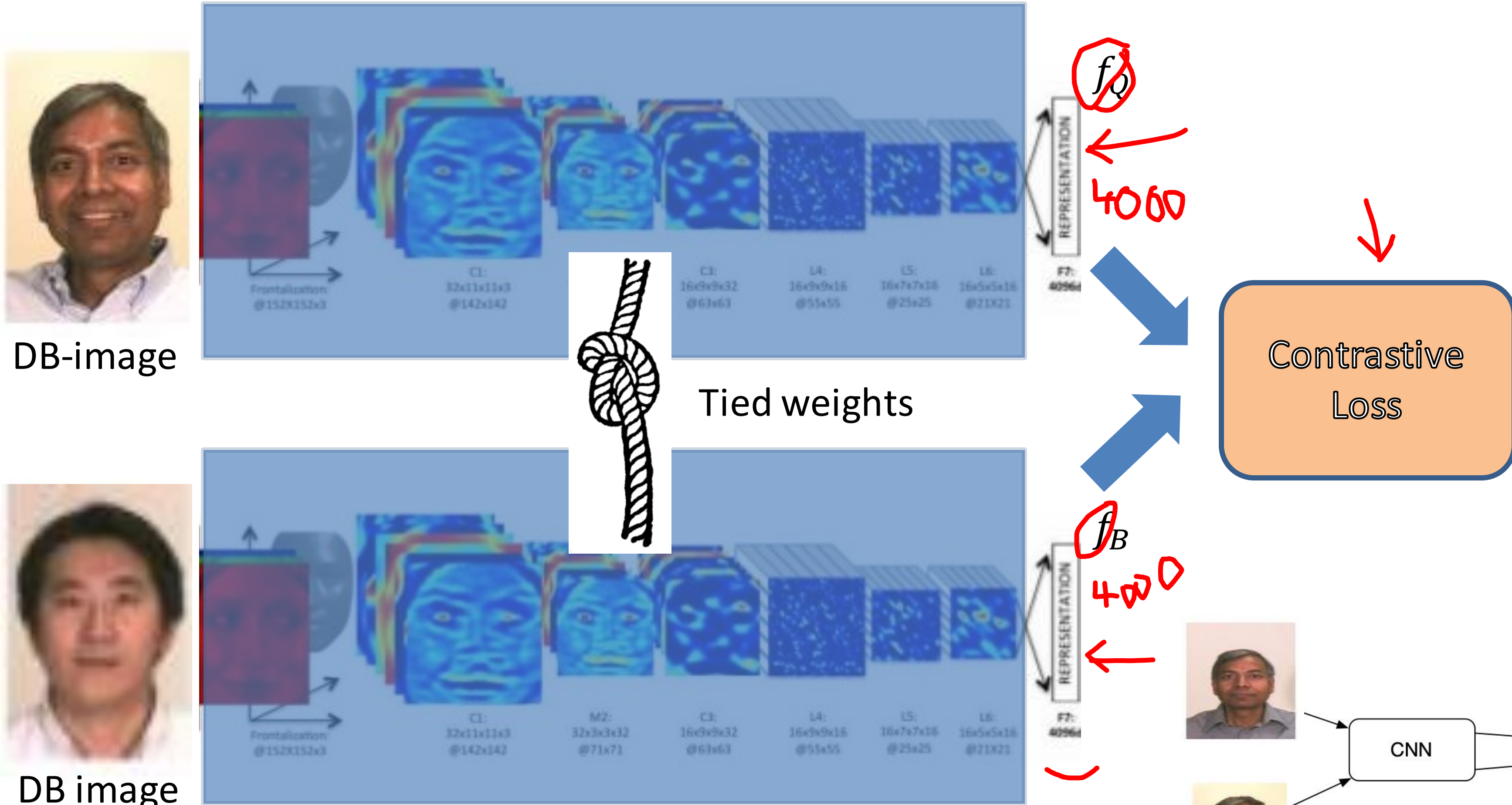


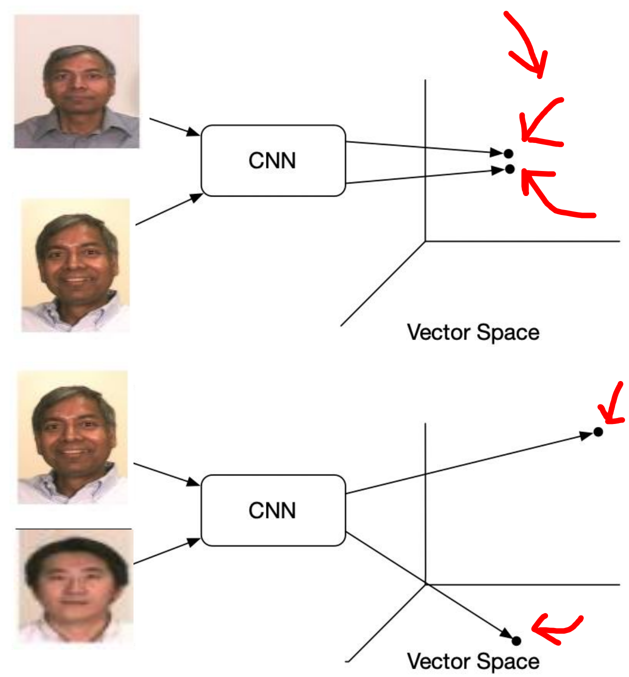
Verification: Approach – 1B



W $\qquad \dfrac{G_1 + G_2}{2}$

Verification: Approach – 1C



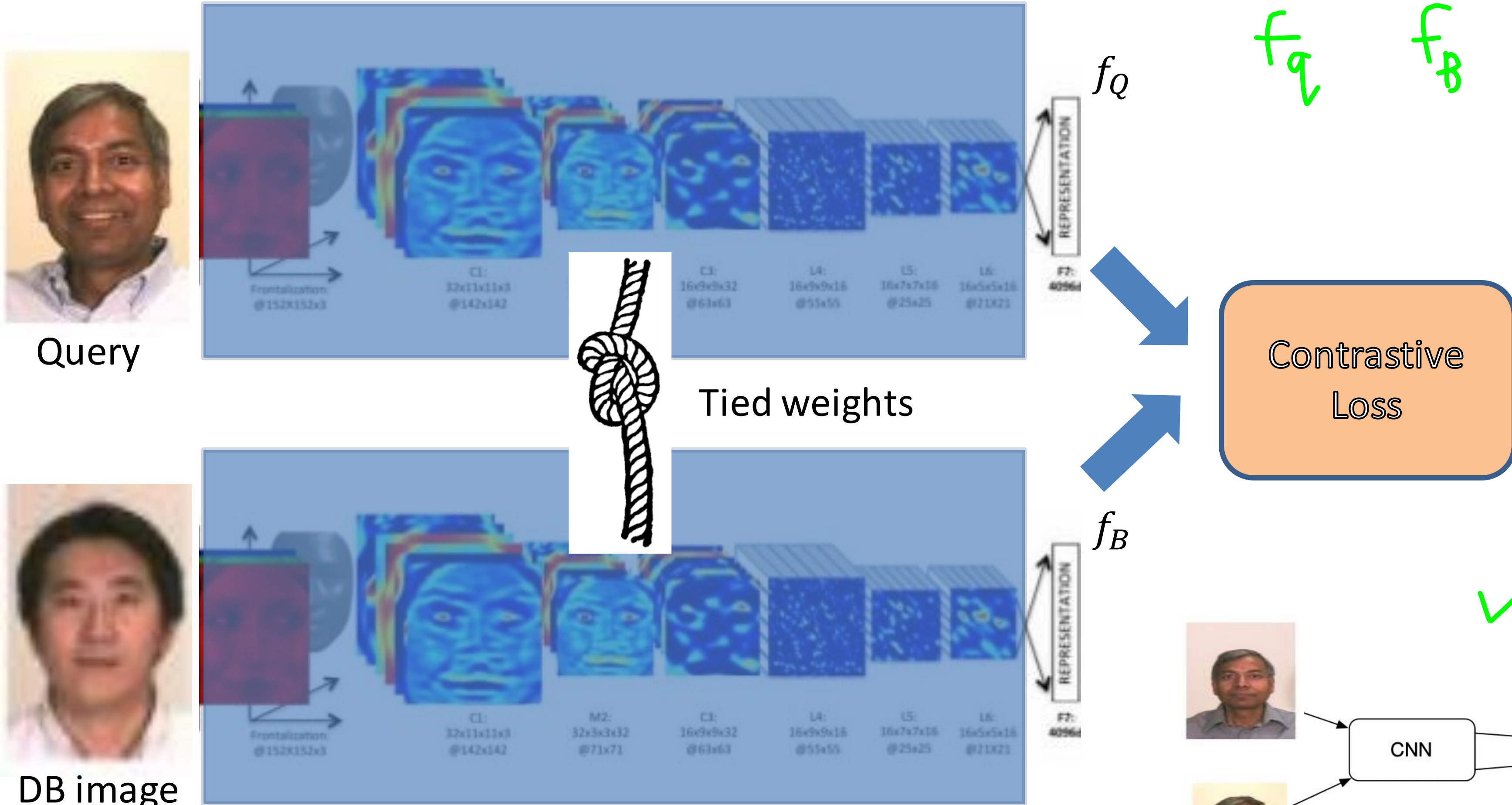


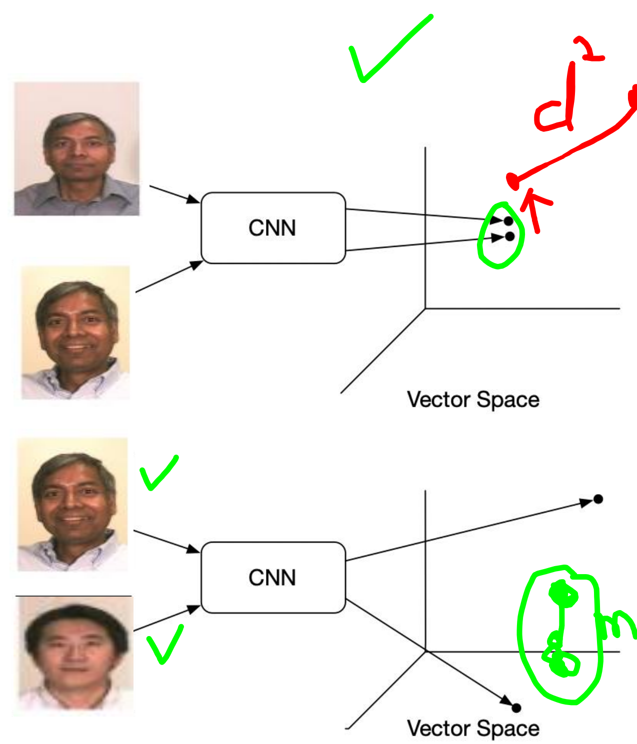
Contrastive Loss:

Learn f_Q , f such that:

- dist(f_Q, f) is large when ids mismatch
- dist(f_Q, f) is small when ids match

Verification: Approach – 1C





Contrastive Loss: $yd^2 + (1-y)\max(margin - d, 0)^2$ >0

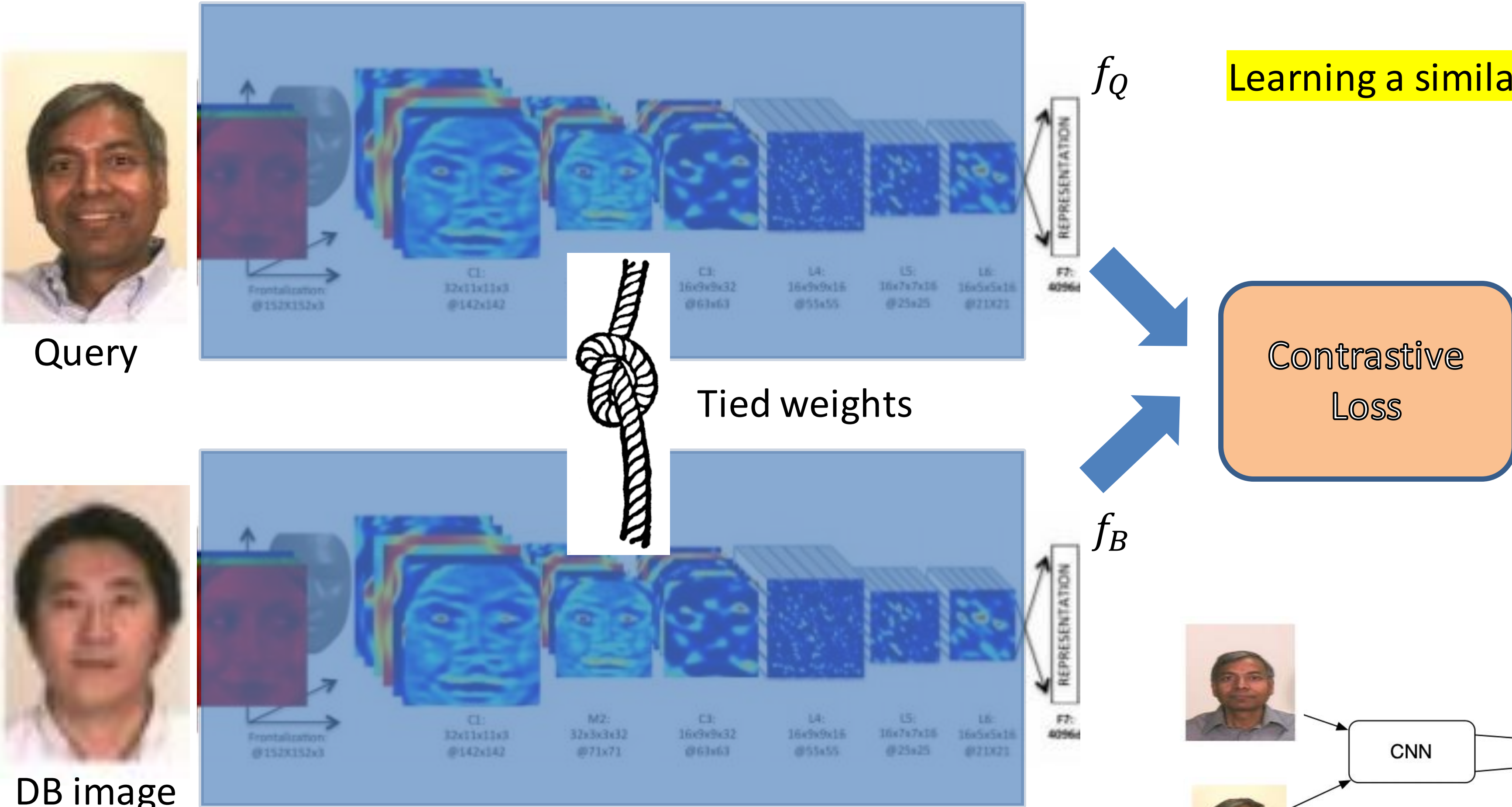
$d<m$

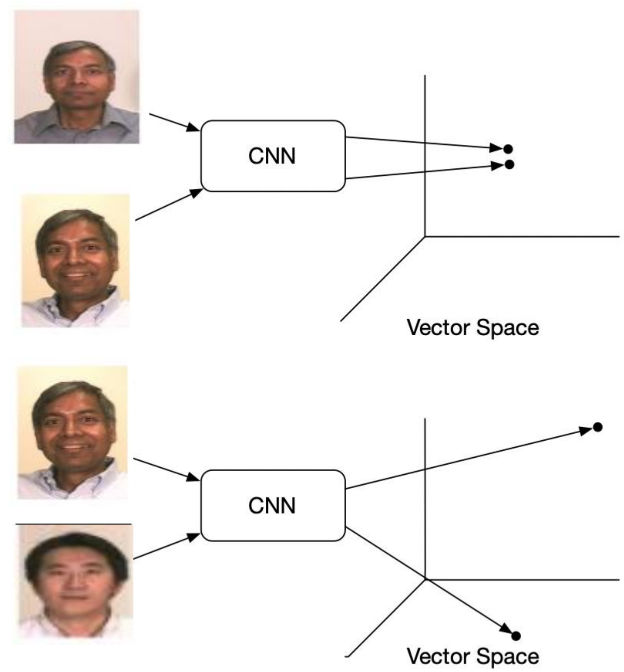
Learn f_Q , f such that:

- d =dist(f_Q, f) is large when ids mismatch (y=0)
- d = dist(f_Q, f) is small when ids match (y=1)

Verification: Approach – 1C

Learning a similarity function





Contrastive Loss: $yd^2 + (1-y)\max(margin - d, 0)^2$

Learn f_Q , f such that:

- d =dist(f_Q, f) is large when ids mismatch (y=0)
- d = dist(f_Q, f) is small when ids match (y=1)

Verification Approach 2

MATCH (1)

3

DB image

3

Query

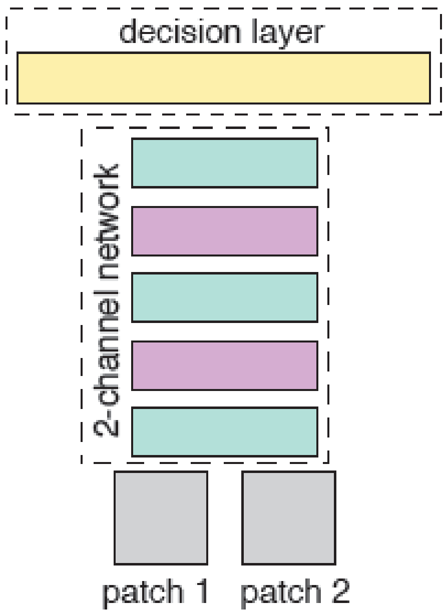
MISMATCH (0)

DB image

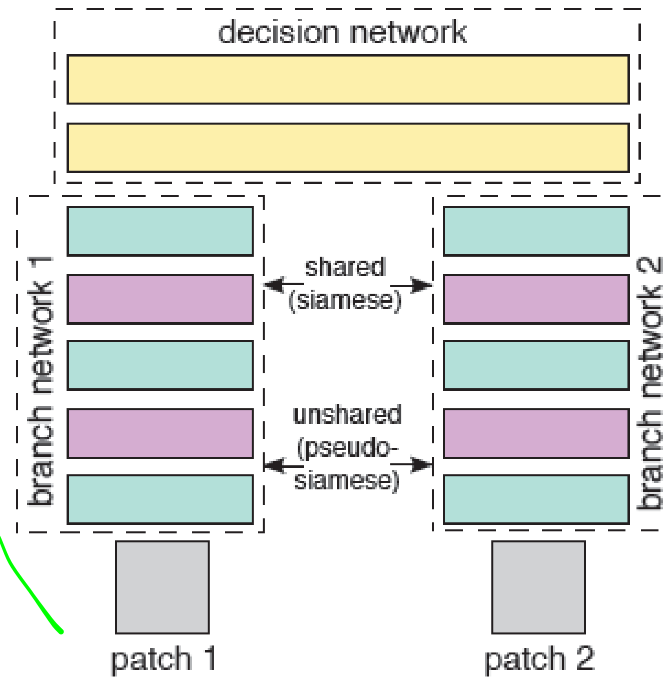
Query

Popular Architecture Varieties

- No one “architecture” fits all!
- Design largely governed by what performs well empirically on the task at hand.



Inputs are merged right at the onset



Inputs are first embedded independently, then merged.

Zagoruyko, S. and Komodakis, N., 2015. Learning to compare image patches via convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4353-4361).

Siamese Network

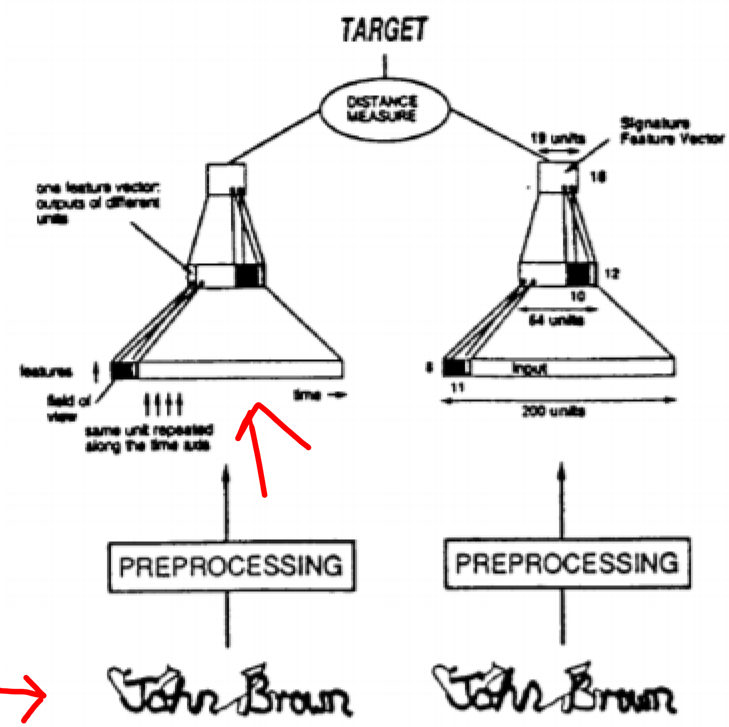
Application in Signature Verification

- The input is 8(feature) x 200(time) units.

- The cosine distance was used, (1 for genuine pairs, -1 for forgery pairs)

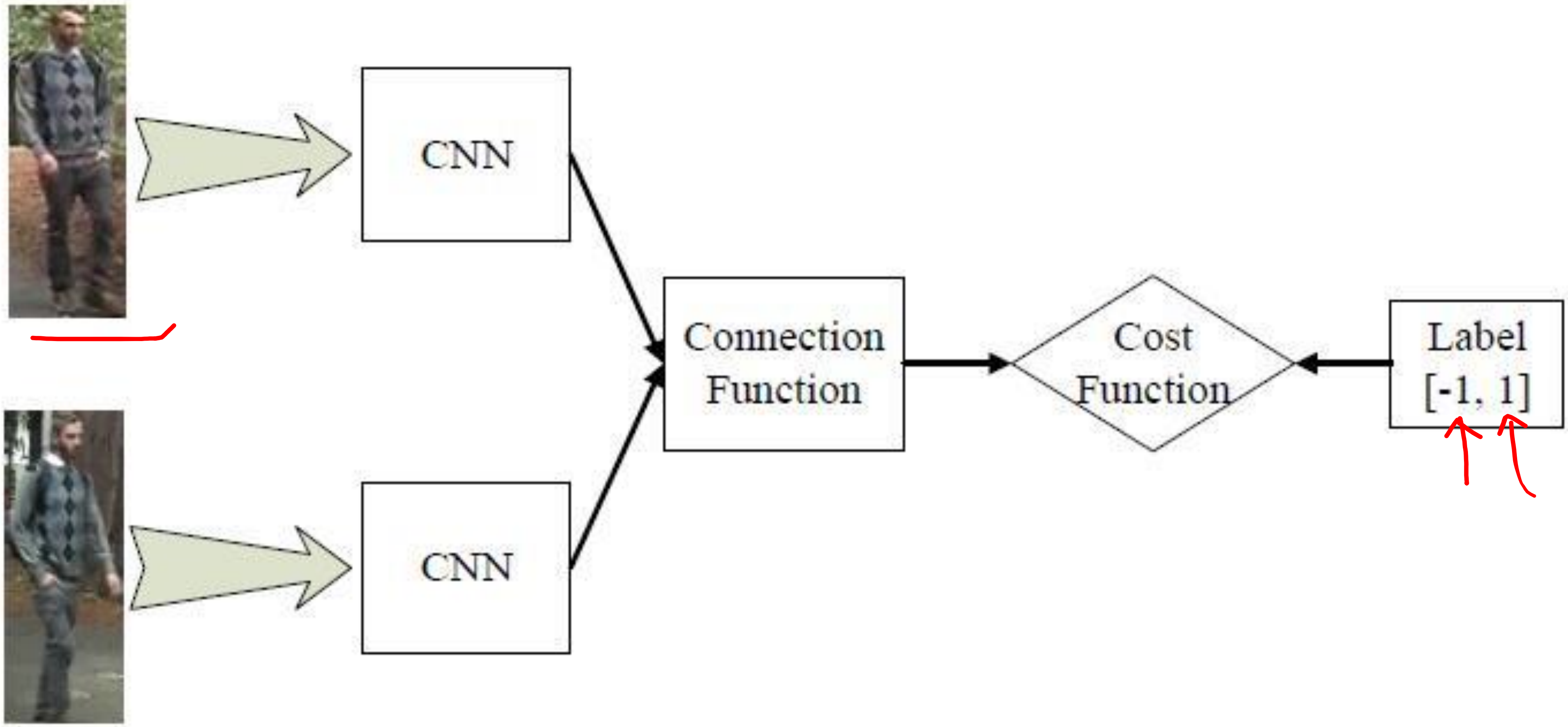
$\cos(f_1, f_2)$

-1 +1



Bromley J, Guyon I, Lecun Y, et al. Signature Verification using a" Siamese" Time Delay Neural Network, NIPS Proc. 1994

Siamese Network (Person re-id)

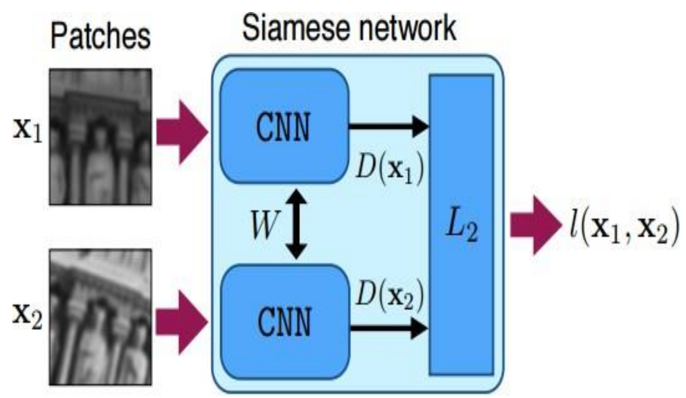


http://www.fubin.org/research/Person_ReID/Person_ReID.html

Siamese CNN – Training

- Update each of the two streams independently and then average the weights.

$\partial l / \partial D(x_1)$

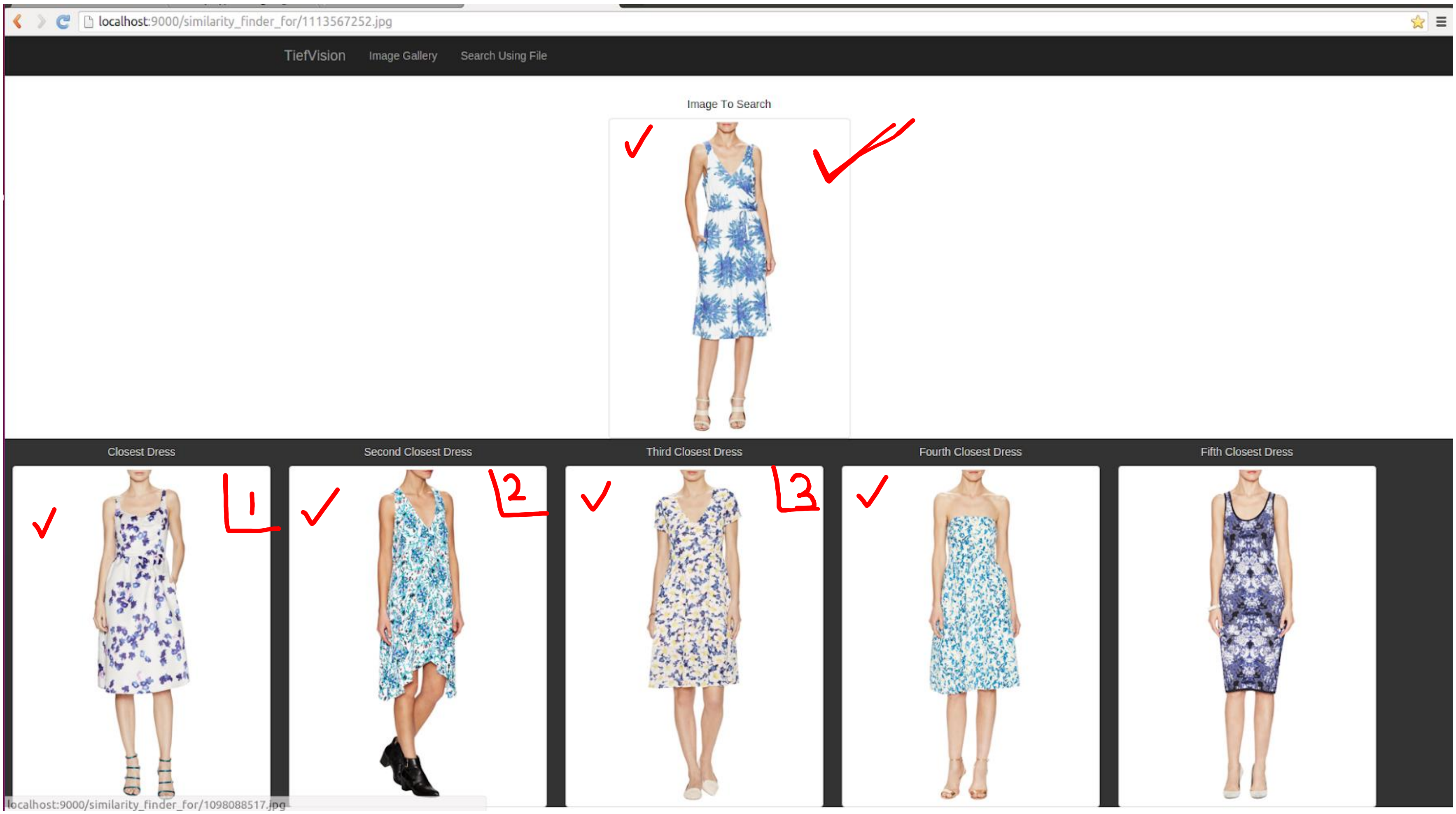


$\partial l / \partial D(x_2)$

Applications

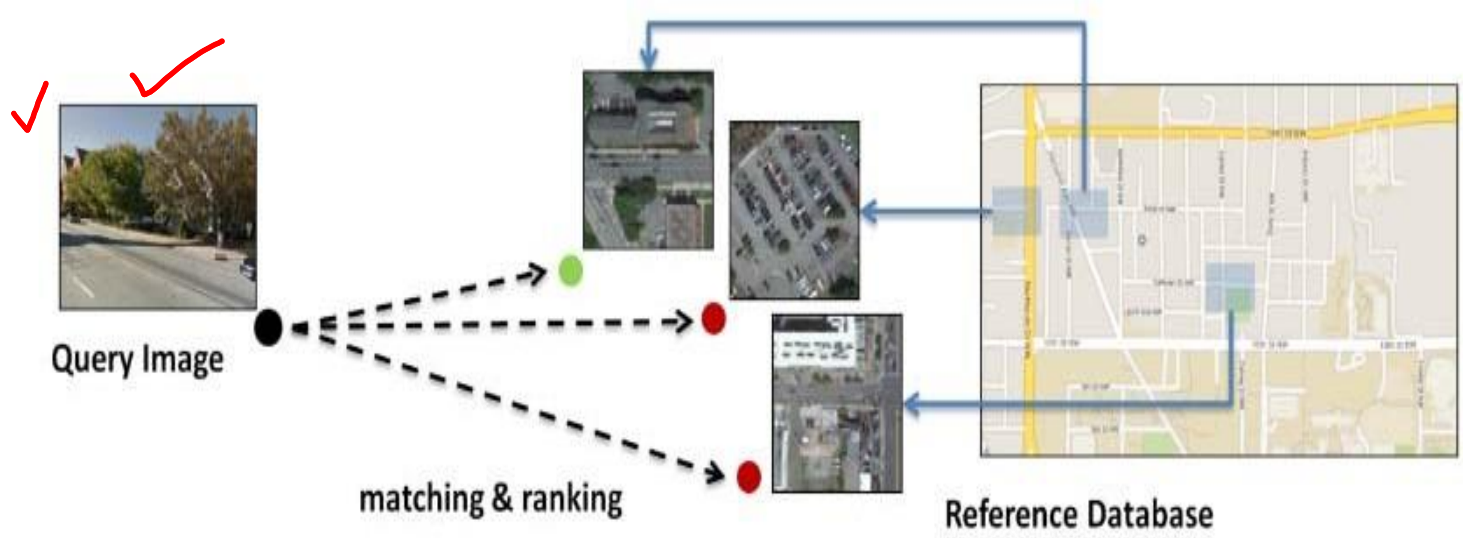
Retrieval

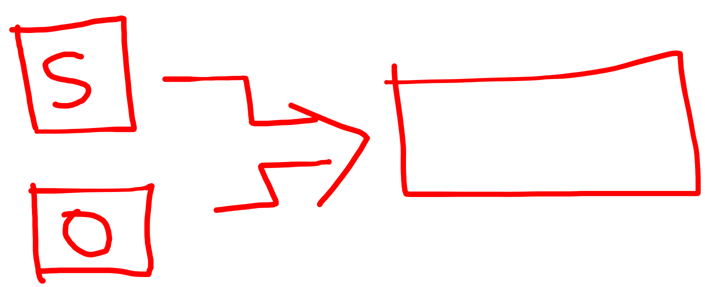
Ranking



https://github.com/paucarre/tiefvision

Street-View to Overhead-View Image Matching

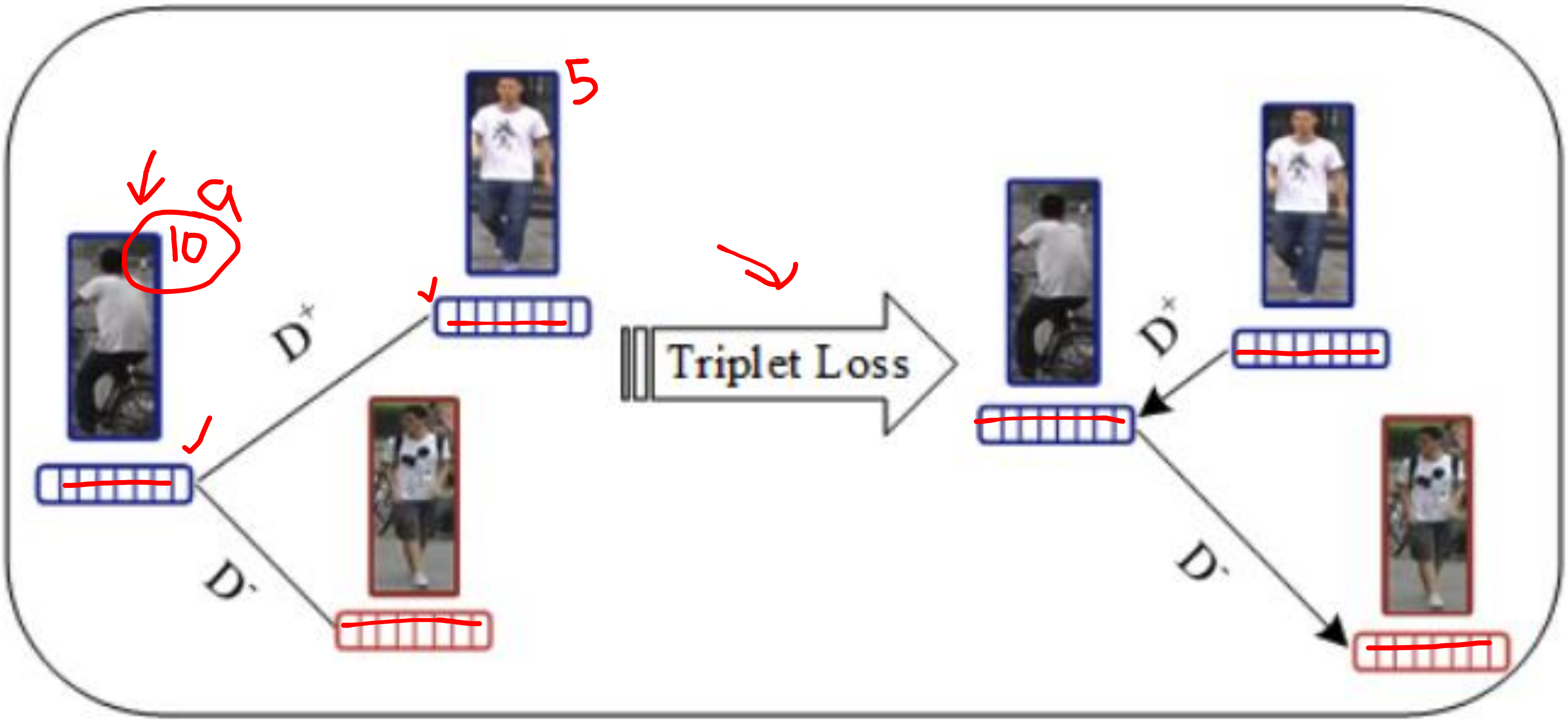




Vo, N.N. and Hays, J., 2016, October. Localizing and orienting street views using overhead imagery. In European Conference on Computer Vision (pp. 494-509).

Many variants exist

- Popular Loss Function – Triplet Loss



Code Reference

https://medium.com/@prabhnoor0212/siamese-network-keras-31a3a8f37d04

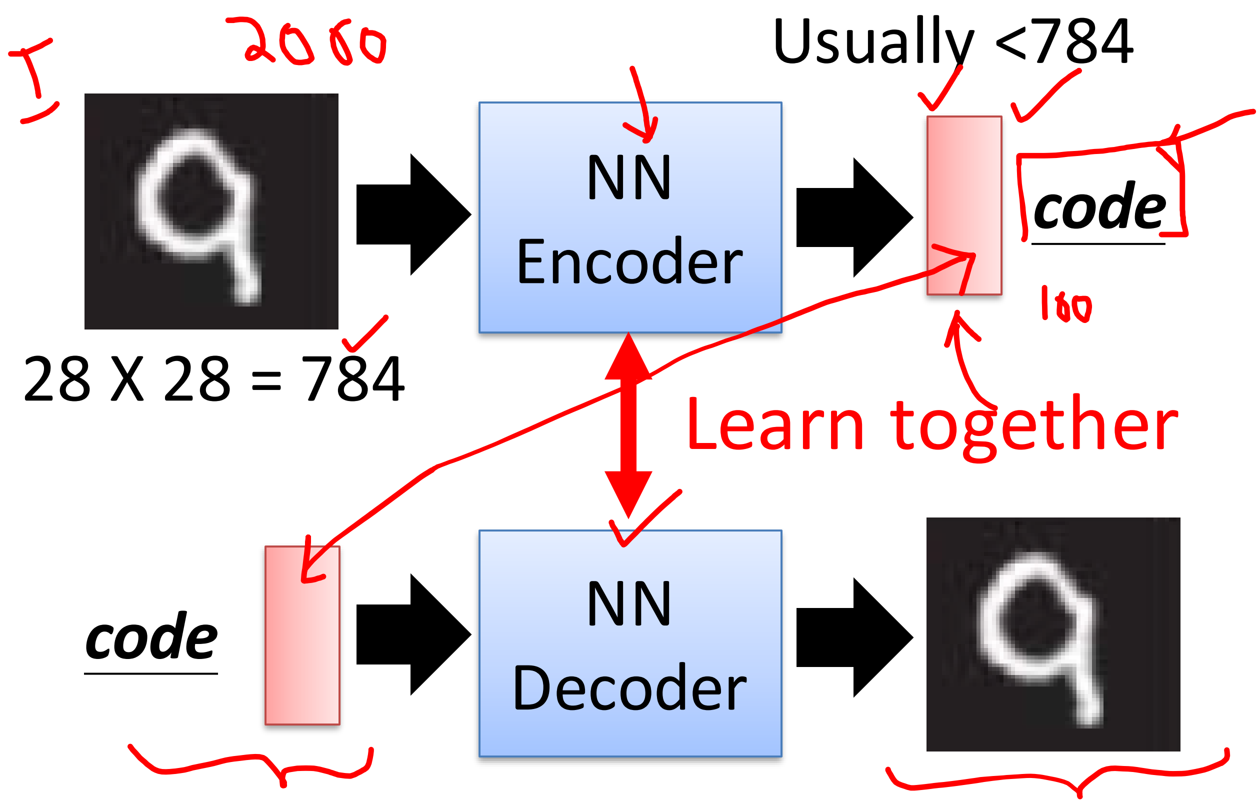
Unsupervised Learning:

Deep Auto-encoder

Unsupervised Learning

"We expect unsupervised learning to become far more important in the longer term. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object."
– LeCun, Bengio, Hinton, Nature 2015

Auto-encoder



Compact representation of the input object

Can reconstruct the original object

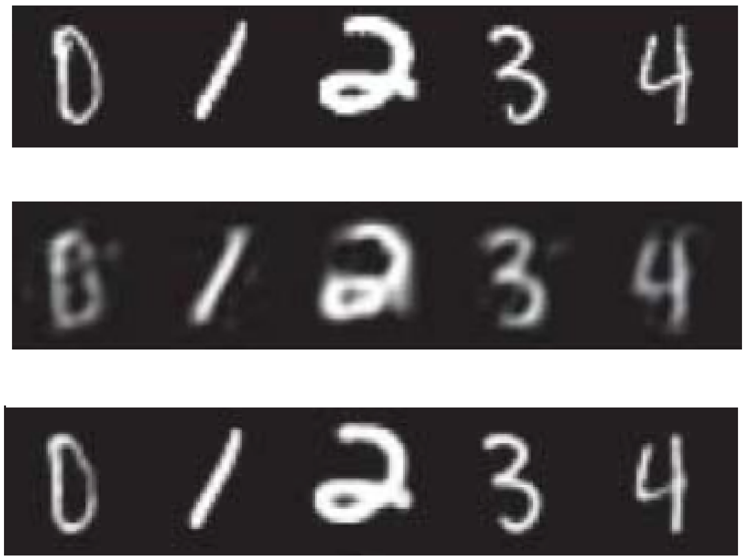
Deep Auto-encoder

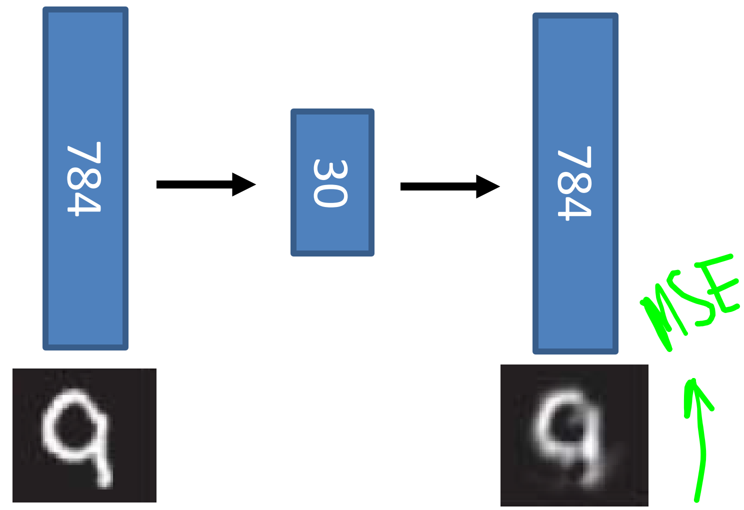
$d(x, \hat{x})$

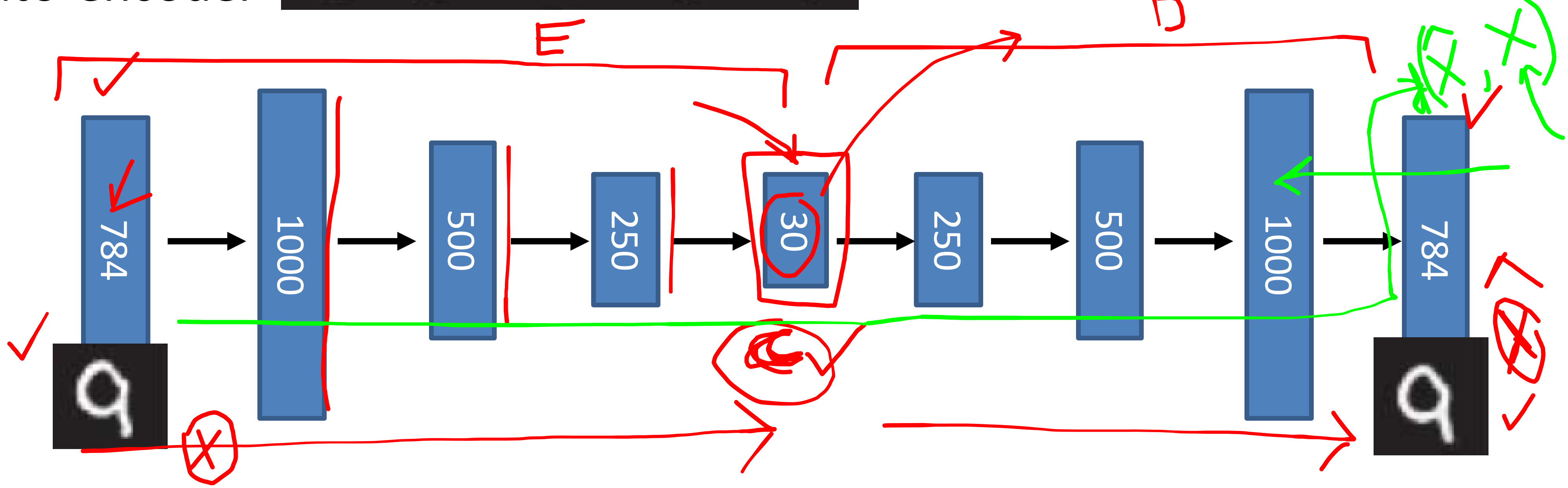
Original Image

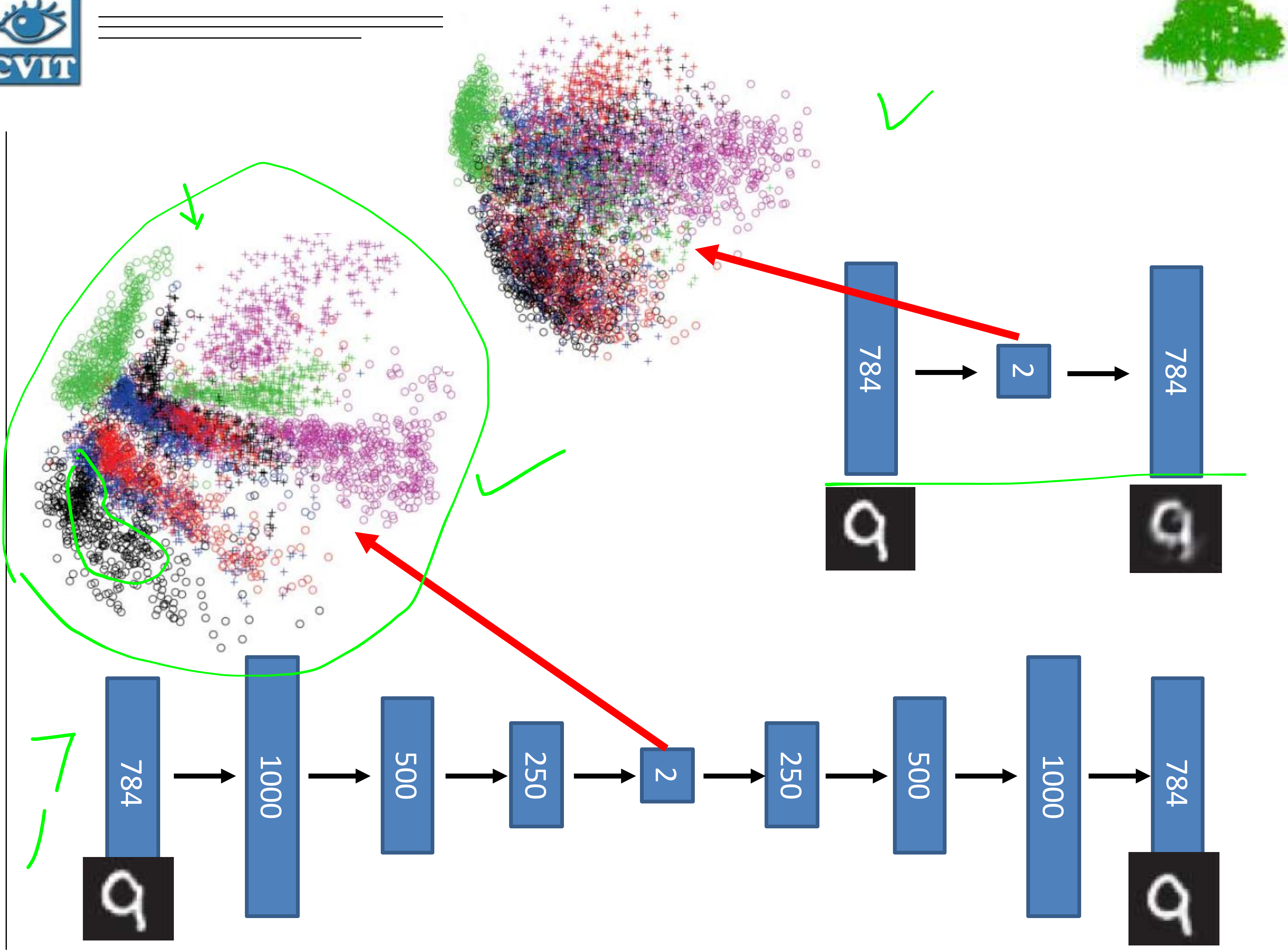
PCA

Deep Auto-encoder



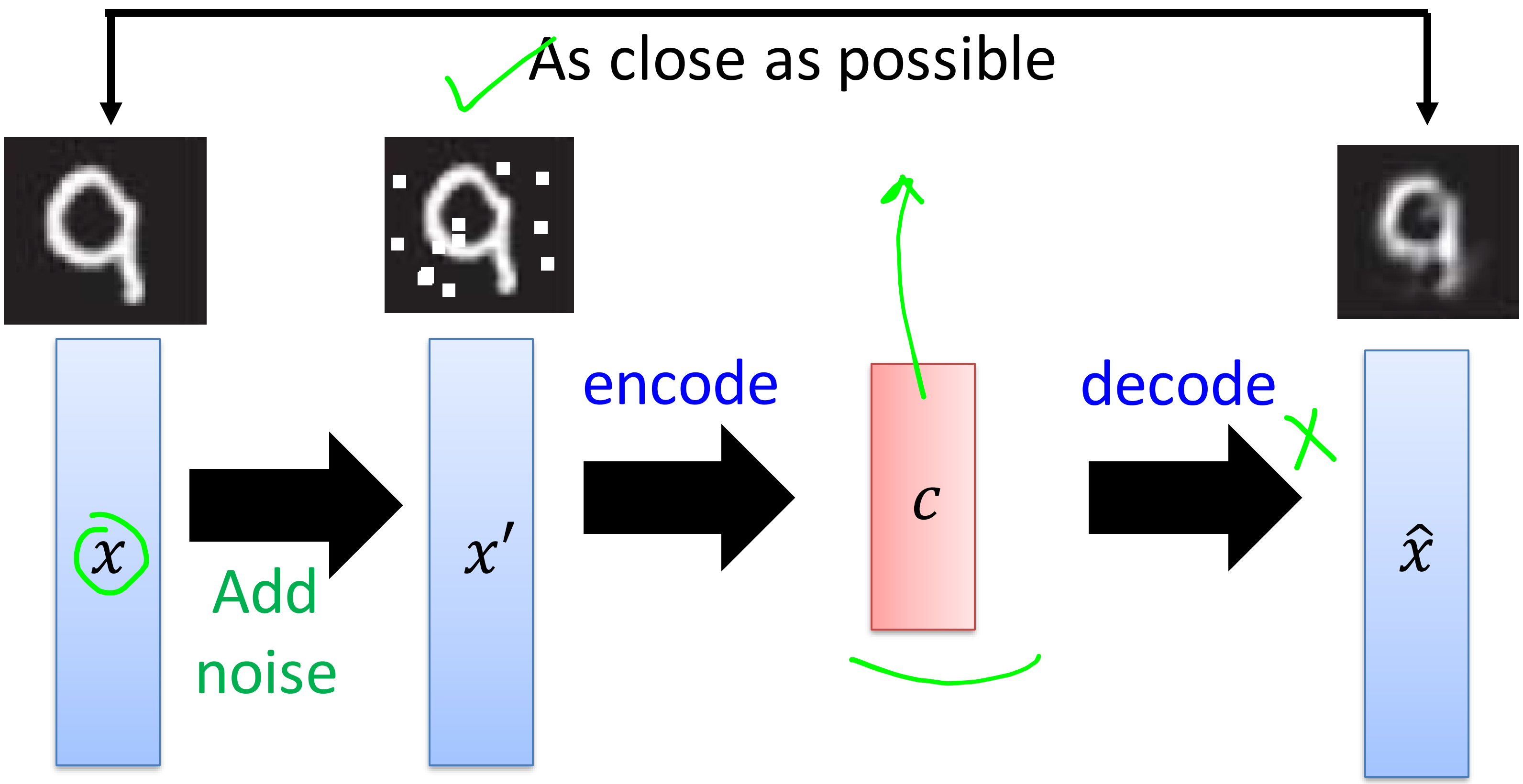





784
2
784
784
1000
500
250
2
250
500
1000
784

Auto-encoder

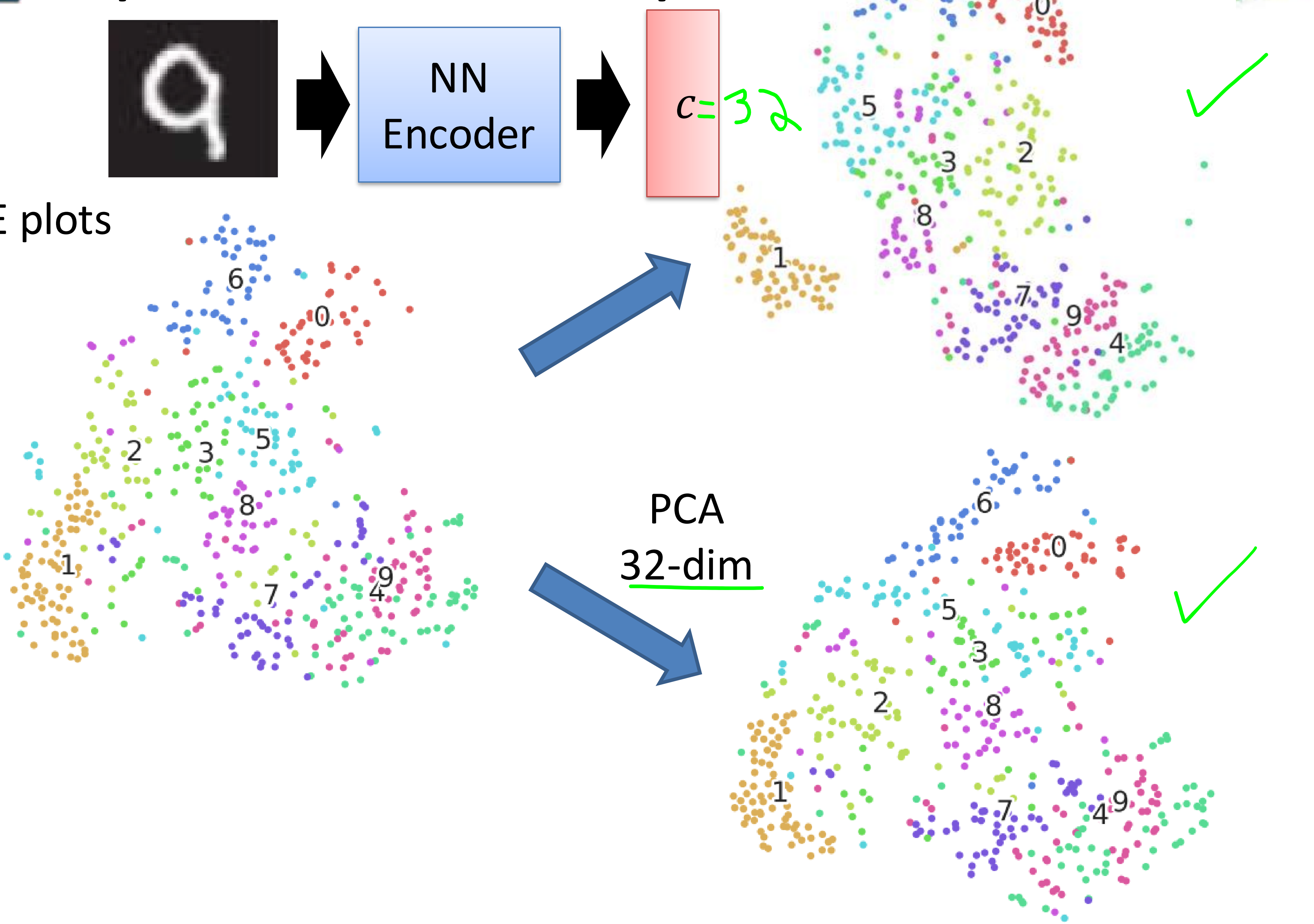
De-noising auto-encoder



Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." *ICML,* 2008.

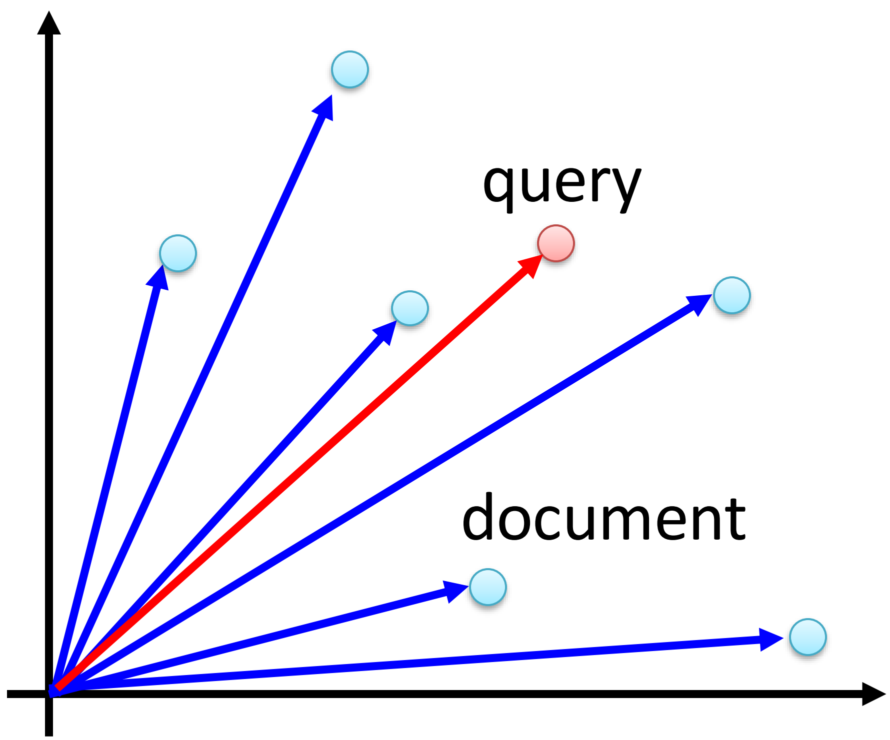
Deep Auto-encoder - Example

tSNE plots



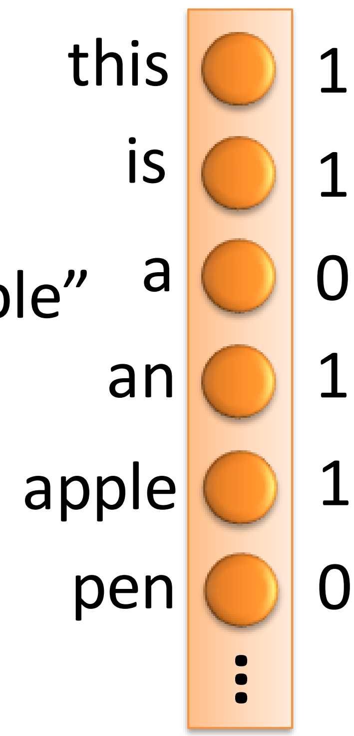
Auto-encoder – Text Retrieval

Vector Space Model



Bag-of-word

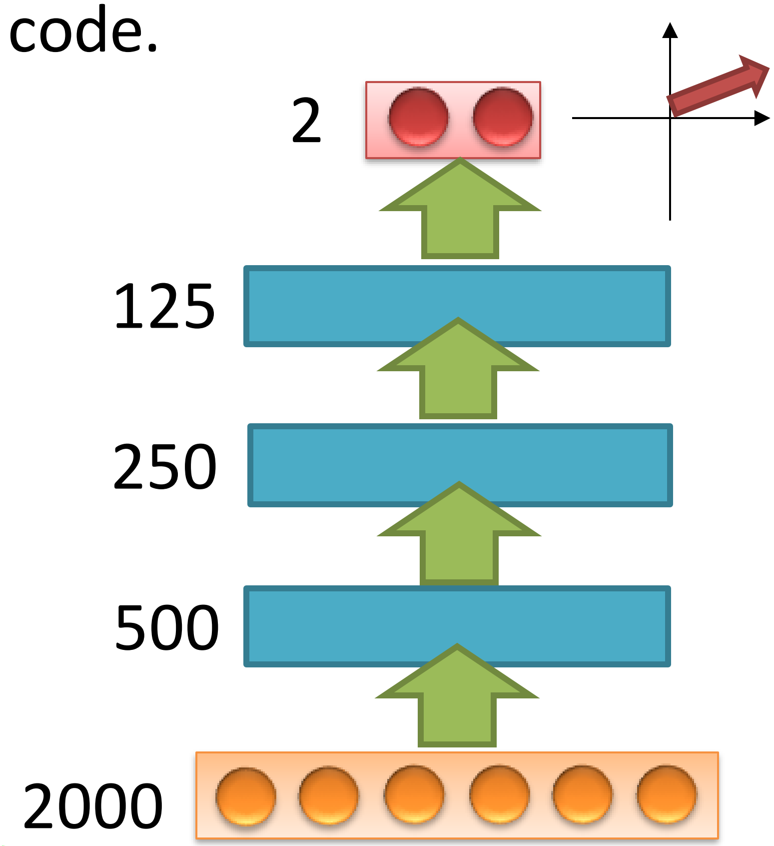
word string:
“This is an apple”



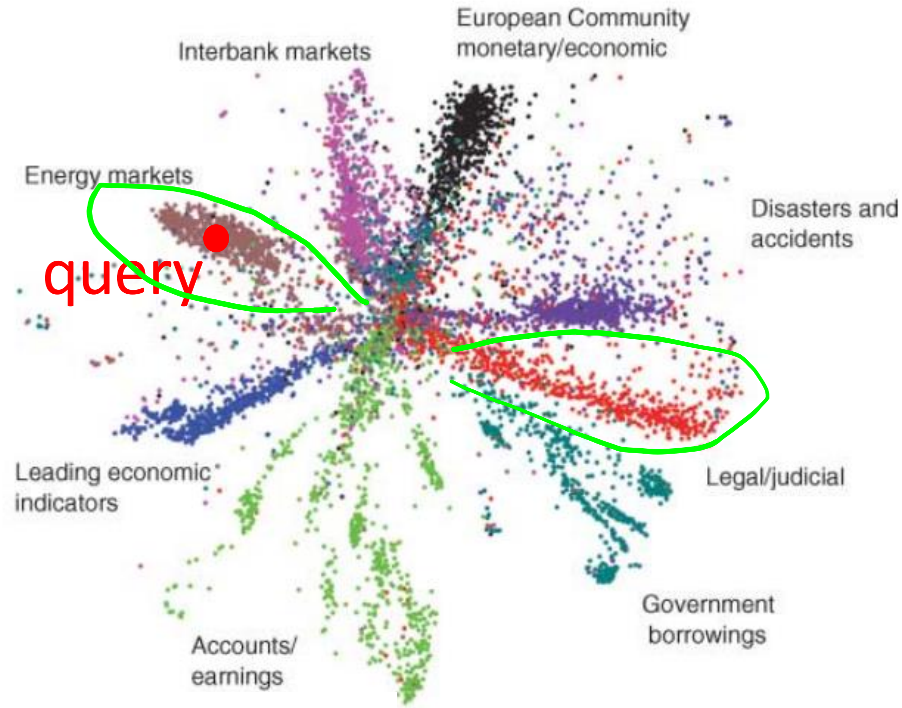
Semantics are not considered.

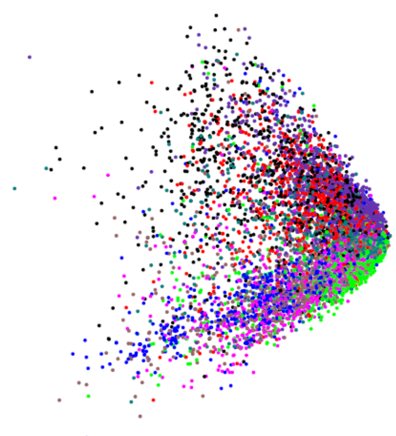
Auto-encoder – Text Retrieval

The documents talking about the same thing will have close code.



Bag-of-word
(document or query)





LSA: project documents to 2 latent topics

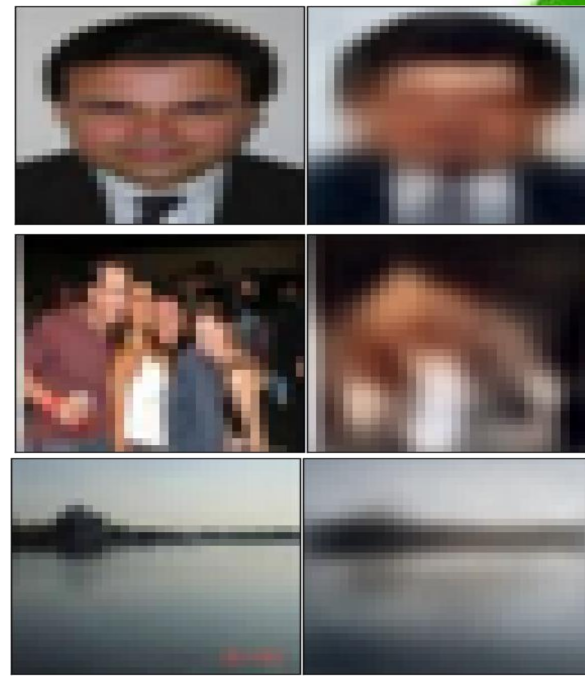
Auto-encoder – Similar Image Search

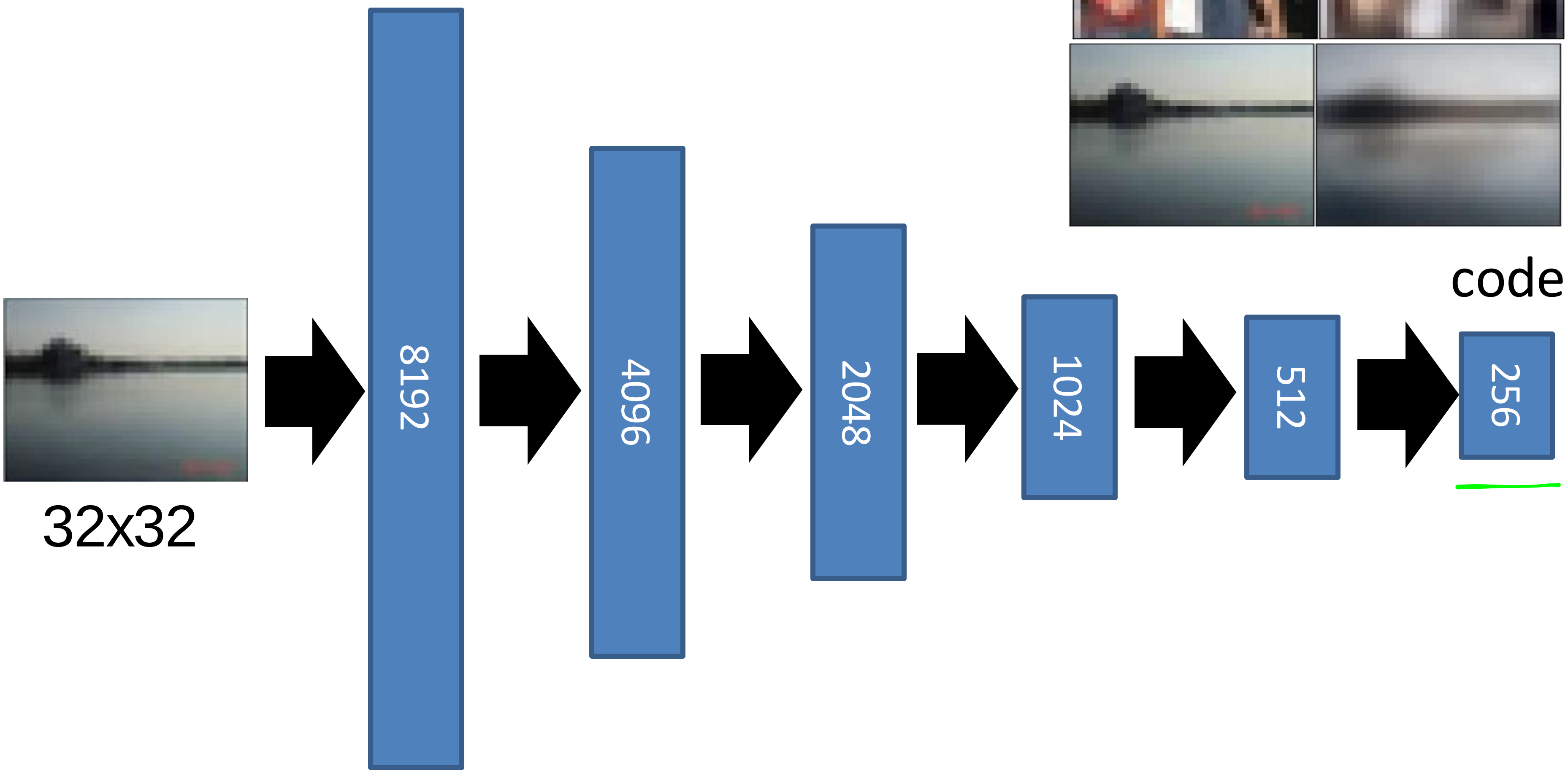
Retrieved using Euclidean distance in pixel intensity space

(Images from Hinton’s slides on Coursera)

Reference: Krizhevsky, Alex, and Geoffrey E. Hinton. "Using very deep autoencoders for content-based image retrieval." *ESANN*. 2011.

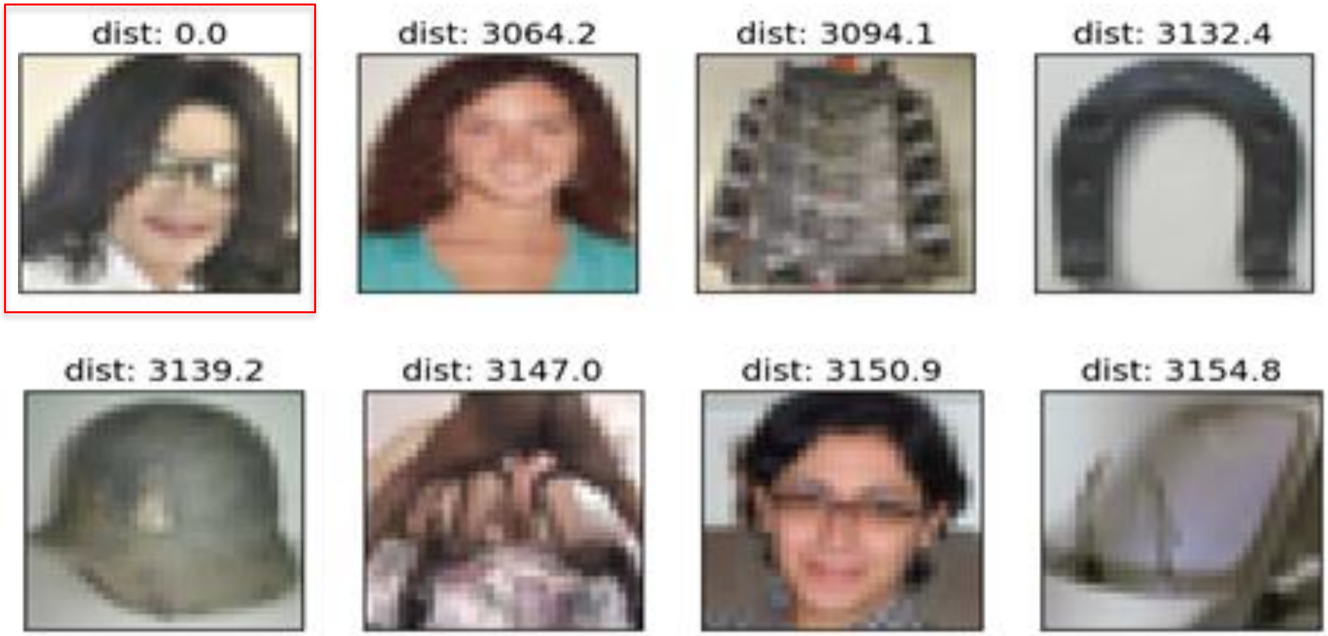
Auto-encoder – Similar Image Search





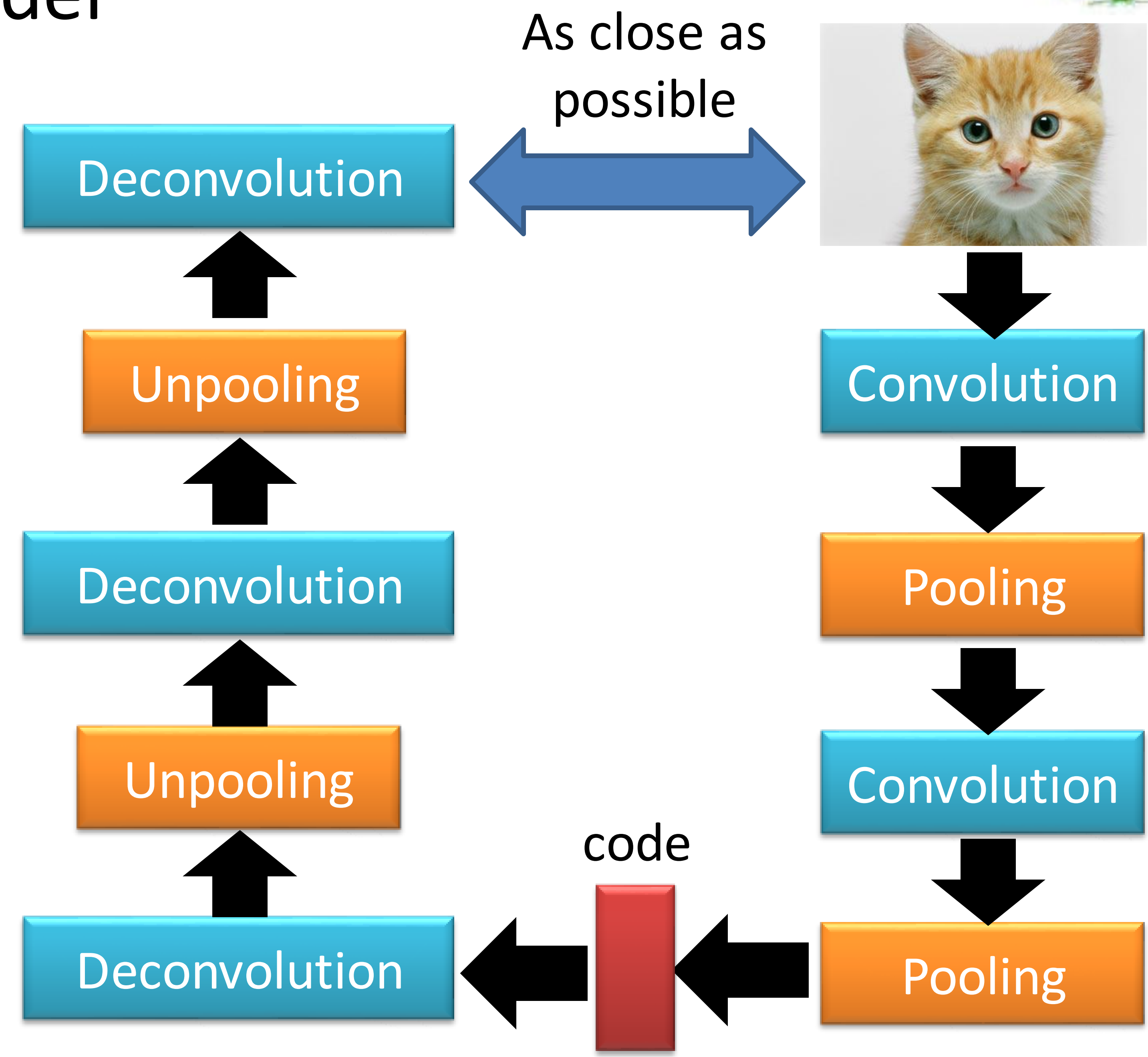
(crawl millions of images from the Internet)

Retrieved using Euclidean distance in pixel intensity space

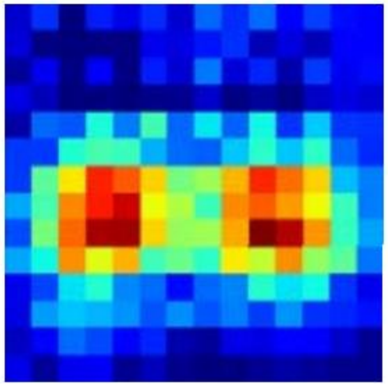


retrieved using 256 codes

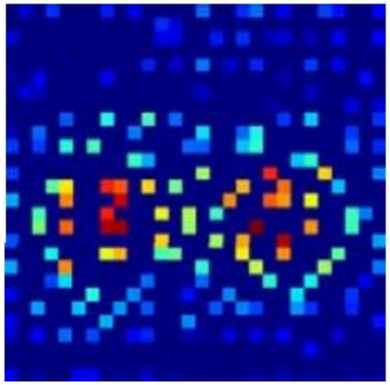
Auto-encoder for CNN



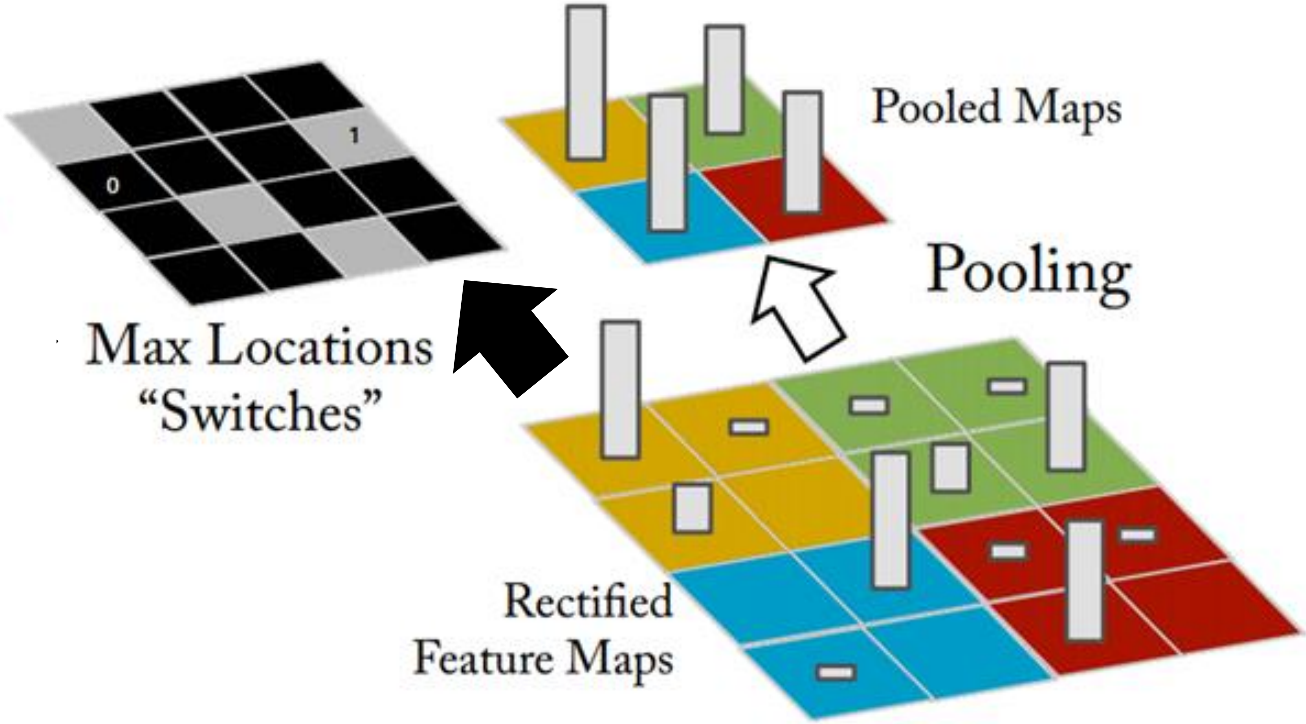
CNN -Unpooling



14 x 14



28 x 28

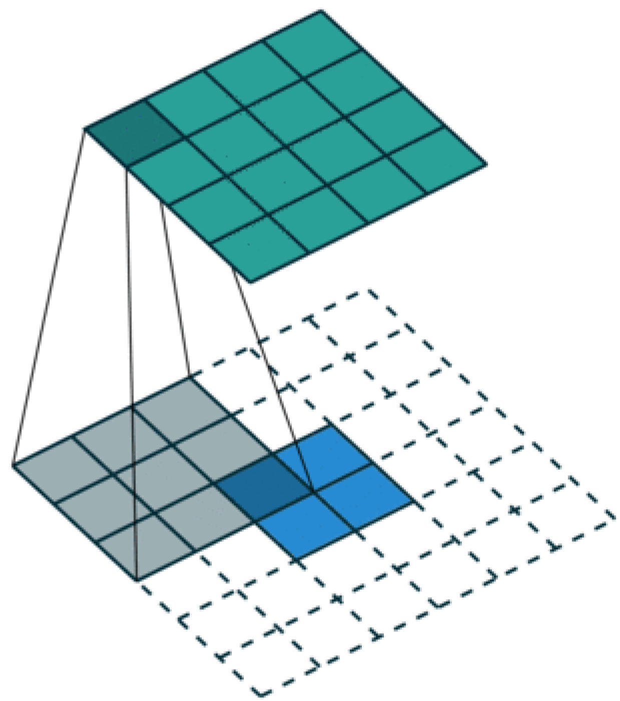


Alternative: simply repeat the values

Source of image : https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/image_segmentation.html

Deconvolution

Actually, deconvolution is convolution.

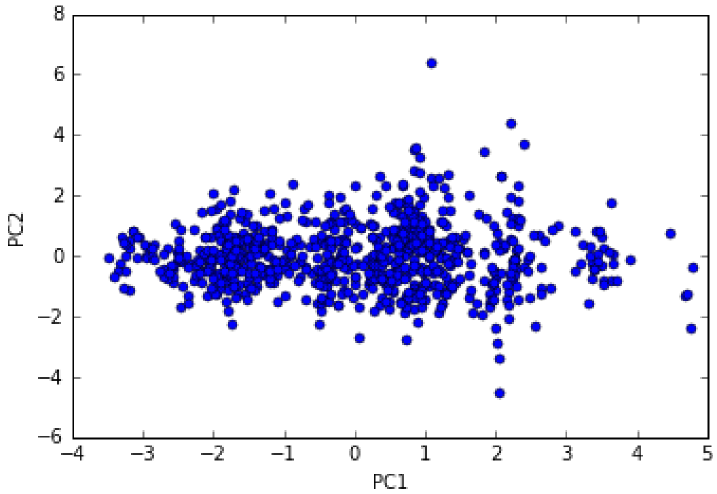


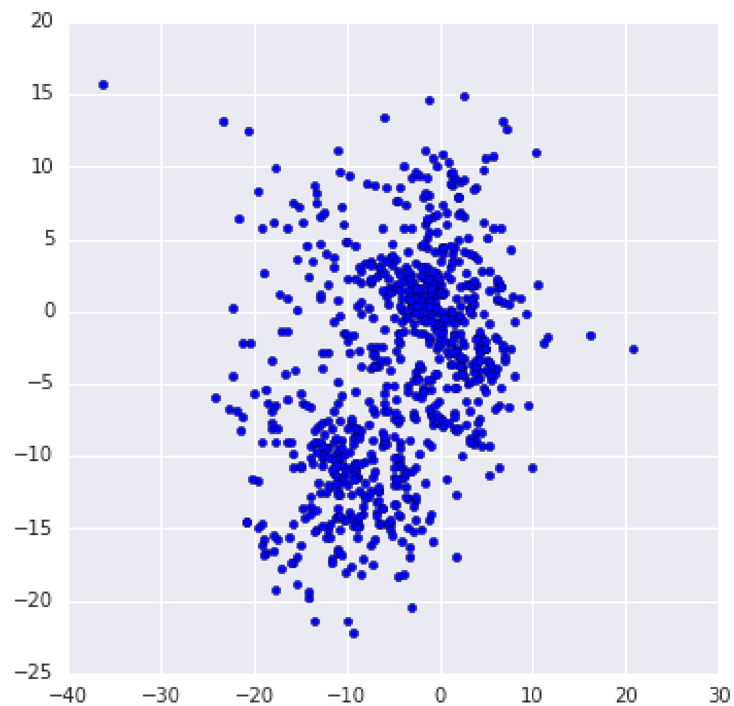
Pokémon

http://140.112.21.35:2880/~tlkagk/pokemon/pca.html
http://140.112.21.35:2880/~tlkagk/pokemon/auto.html

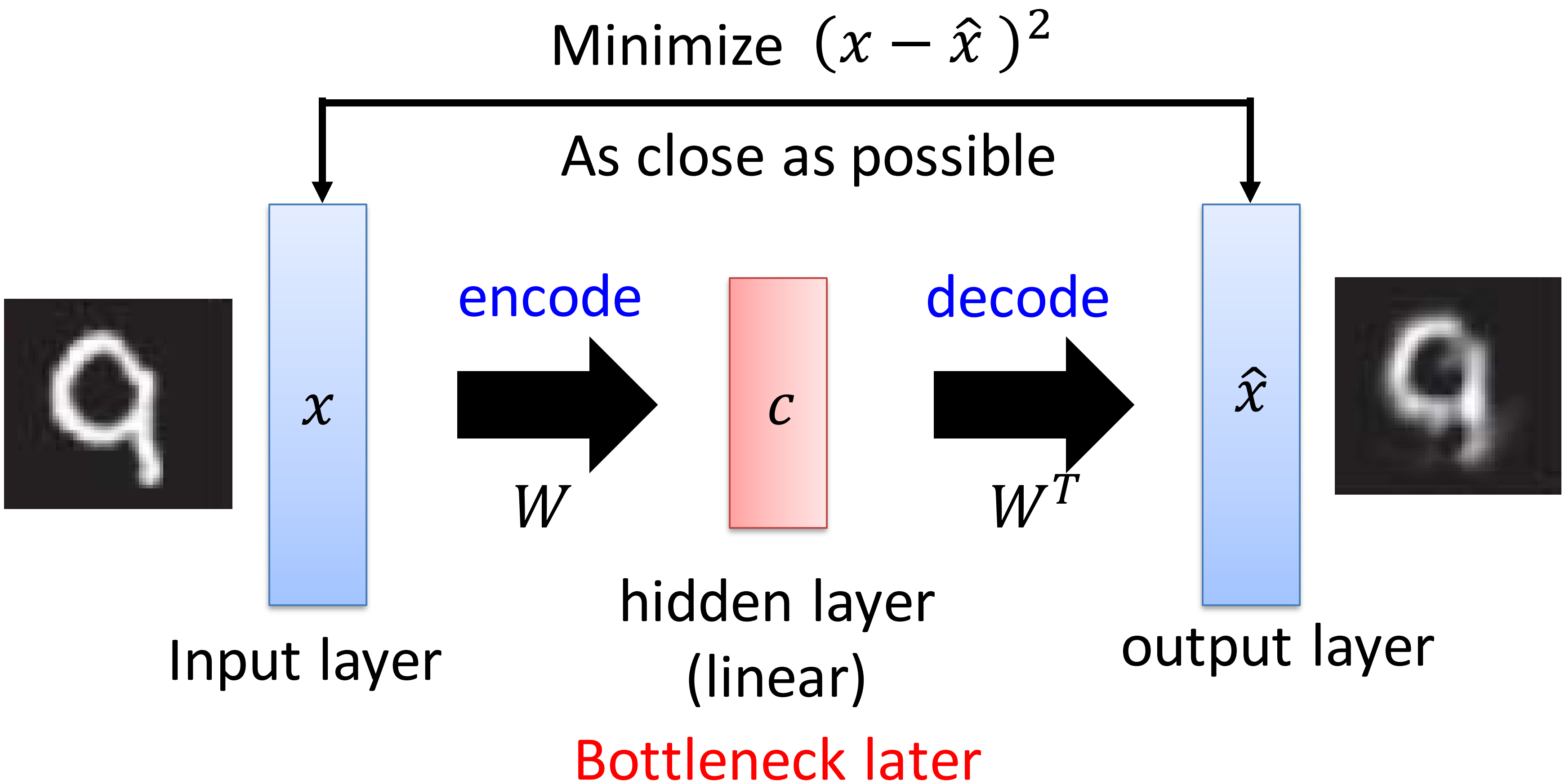
The code is modified from

http://jkunst.com/r/pokemon-visualize-em-all/





PCA ~ Autoencoder with linear layers



Output of the hidden layer is the code

Code Reference

https://blog.keras.io/building-autoencoders-in-keras.html