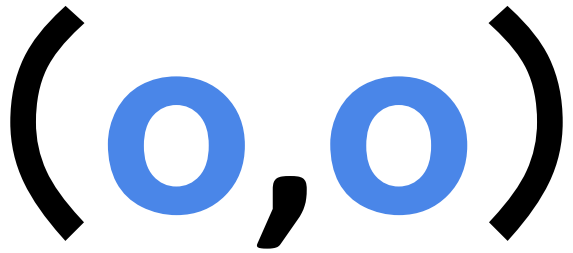


TiefVision: end-to-end image similarity search engine

Pau Carré Cardona (pau dot carre at gmail dot com
)

Demo

TiefVision | https://www.google | TiefVision | TiefVision | TiefVision | TiefVision

localhost:9000/similarity_finder_for/1098532482.jpg

TiefVision Image Gallery Search Using File

Image To Search

Closest Dress

Second Closest Dress

Third Closest Dress

Fourth Closest Dress

Fifth Closest Dress

localhost:9000/similarity_finder_for/1116221894.jpg

TiefVision Image Gallery Search Using File

Image To Search

Closest Dress

Second Closest Dress

Third Closest Dress

Fourth Closest Dress

Fifth Closest Dress

Image To Search

Closest Dress

Second Closest Dress

Third Closest Dress

Fourth Closest Dress

Fifth Closest Dress

Image To Search

Closest Dress

Second Closest Dress

Third Closest Dress

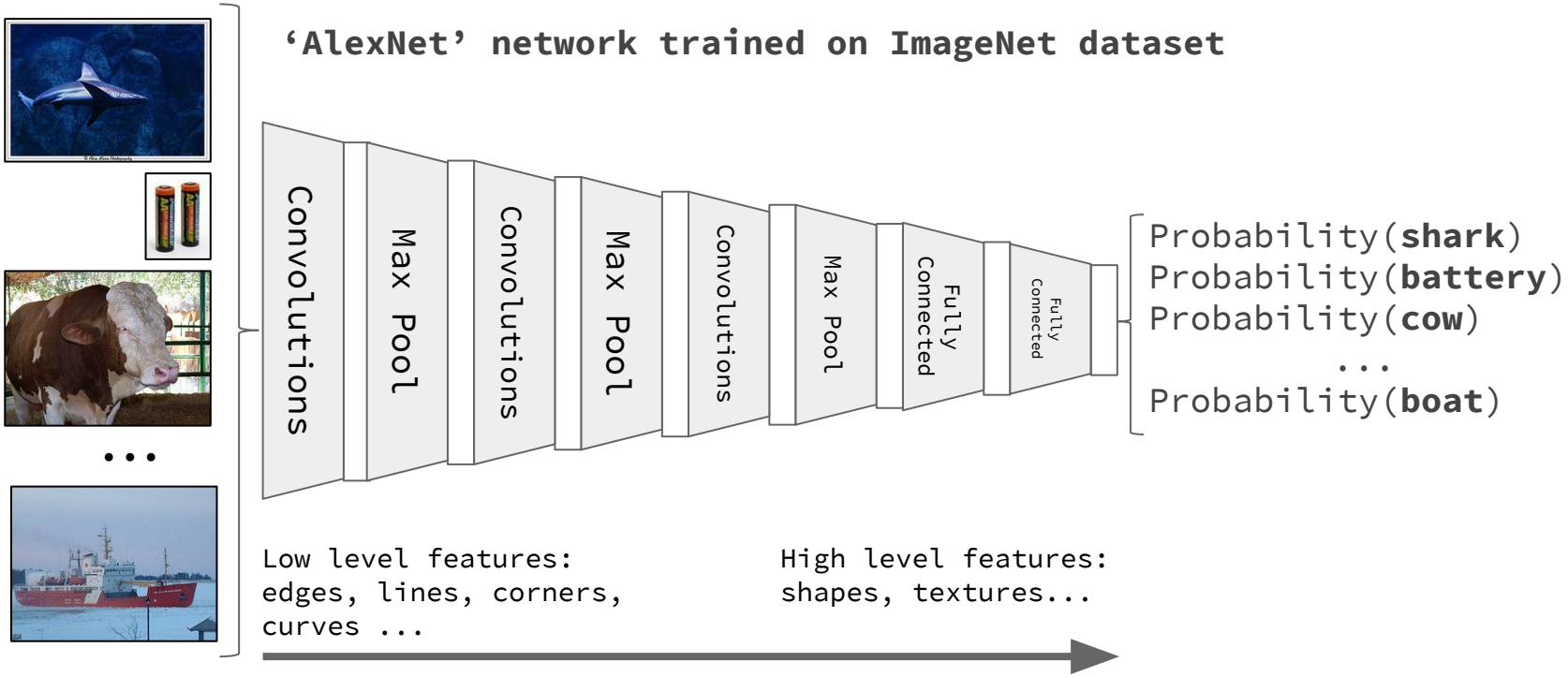
Fourth Closest Dress

Fifth Closest Dress

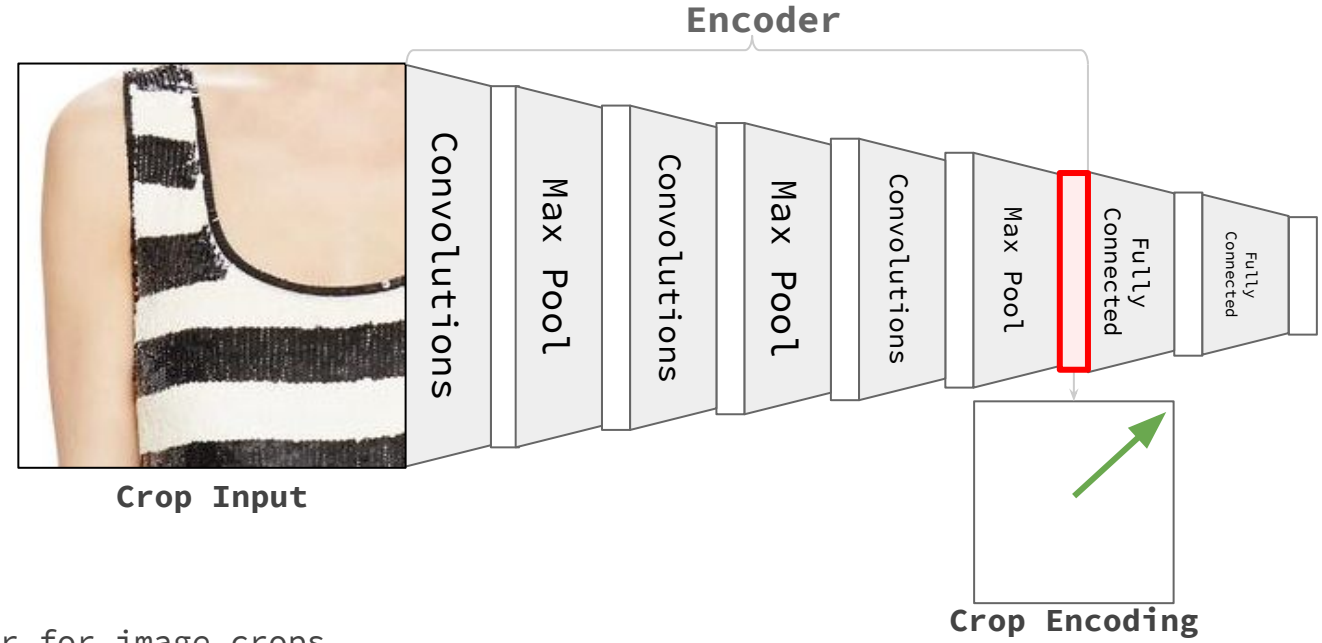
Architecture

1. Dimensionality reduction using transfer learning
2. Image Location (OverFeat)
3. Unsupervised image similarity
4. Supervised image similarity (siamese networks and Deep Rank)

TiefVision: Dimensionality Reduction using Transfer Learning



TiefVision: Dimensionality Reduction using Transfer Learning

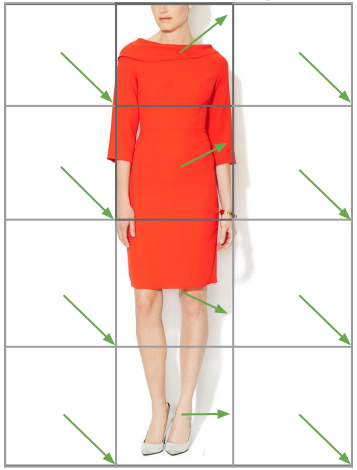


Encoder for image crops

- We need to **reduce the dimensionality of the image crops to be tolerant to small changes and to remove redundant information.**
- For that we can use the **output of the last max pool layer of an existing neural network** such as an AlexNet trained on ImageNet.
- We also reduce the max pool step size to increase the spatial resolution.

TiefVision: Dimensionality Reduction using Transfer Learning

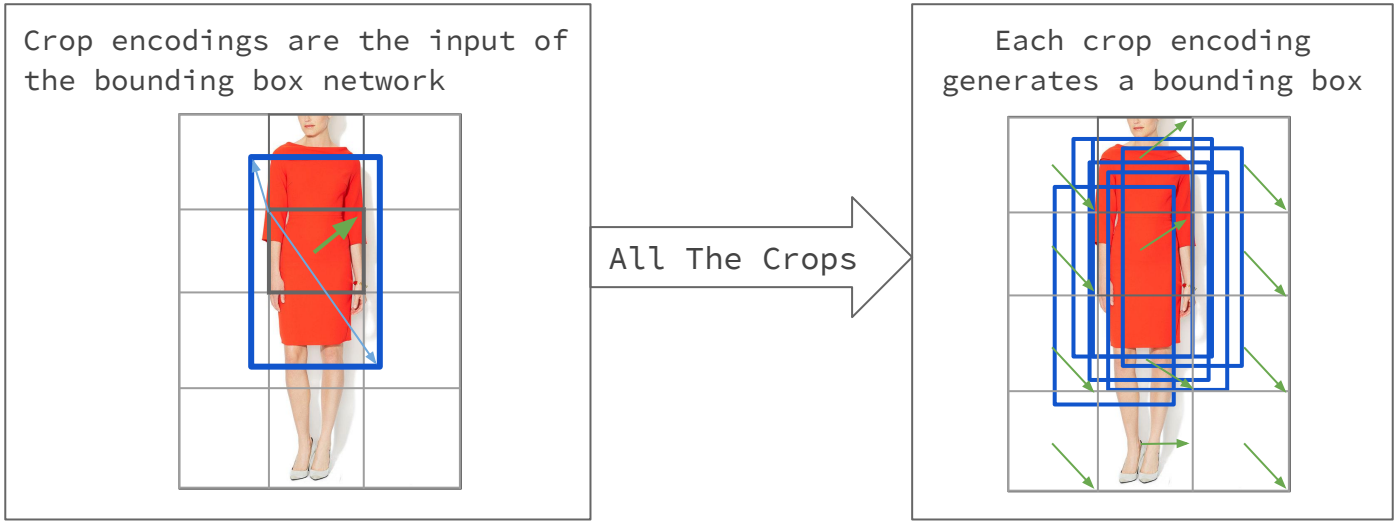
Encodings of all the crops in an image



Encoding

- We convolve the whole image using the convolutional encoder getting encodings for each spatial coordinate.
- If you don't understand how the convolution works, think as if you would make crops and forward them throughout the encoder.

TiefVision: Image location using OverFeat

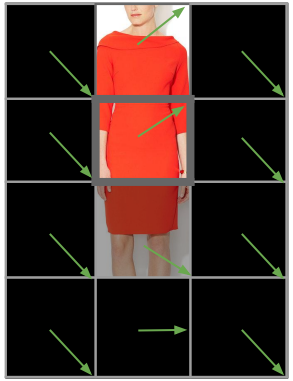


Bounding box regression

- Make crops of the images in such a way they contain a dress in at least 50% of their area.
- Generate input data by encoding the crops using the encoder
- Train a regression network to predict the two 2D relative bounding box points: upper-left point and lower-right point (TiefVision actually uses four neural networks, one for each 1D point).

TiefVision: Image location using OverFeat

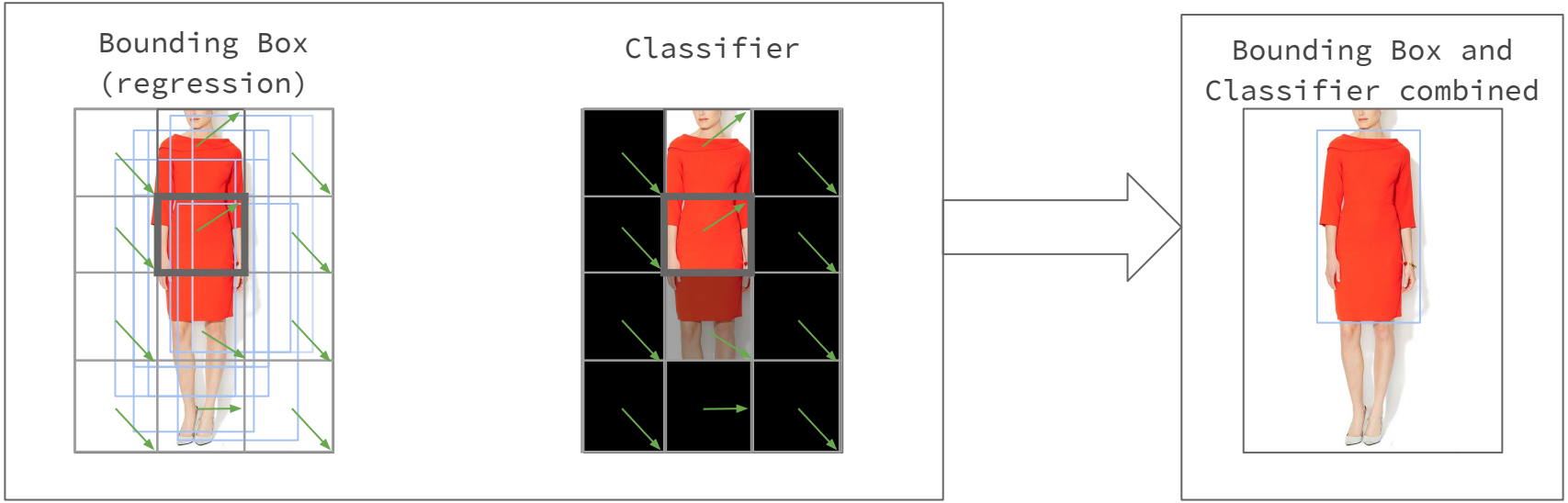
Probability of each crop belonging to a dress



Dress classifier

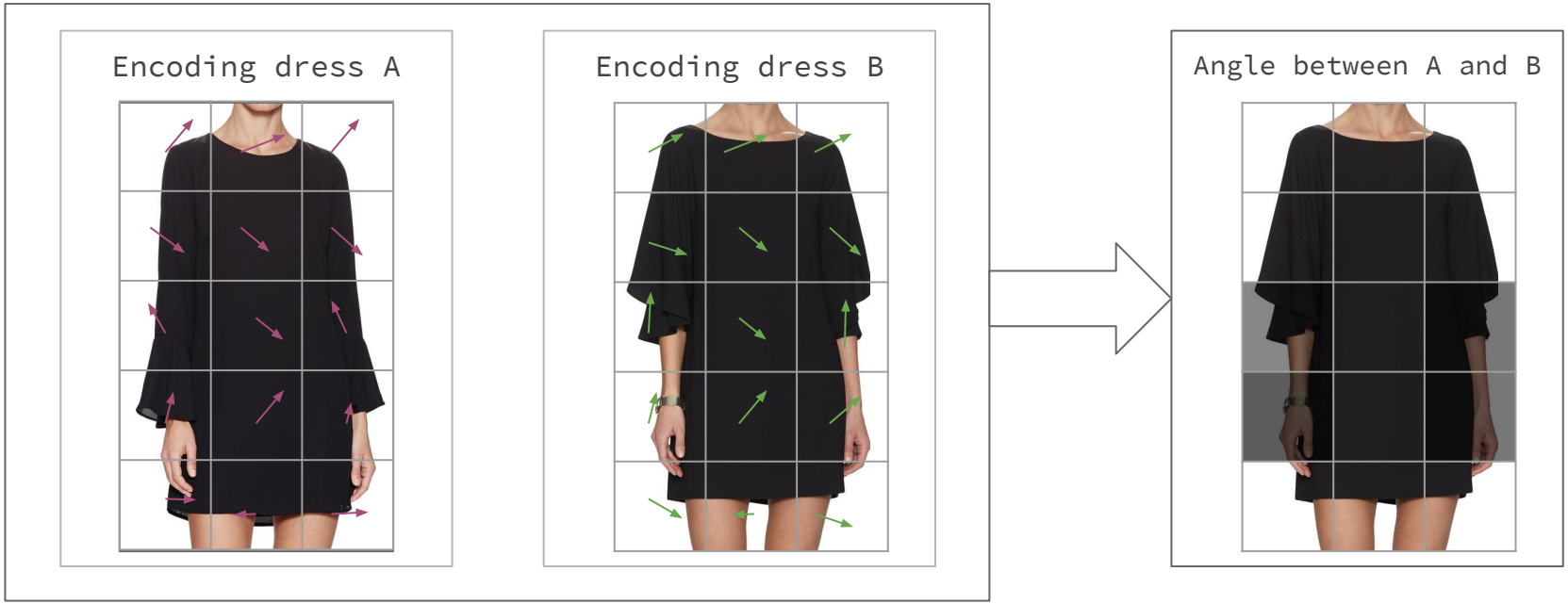
- Generate two classes/types of crops:
  - Crops that contain a dress
  - Crops that don't contain any dress
- Generate input data by encoding all the crops using the encoder
- Train a 'fully connected' network classifier to predict whether a crop belongs to a dress or not.

TiefVision: Image location using OverFeat



1. Get all the bounding boxes for each encoded crop using the regression network
2. Get the probability of each encoded crop to contain a dress
3. Discard the bounding boxes from encoded crops that don't contain a dress (e.g. probability(dress) < 0.8)
4. Average the resulting bounding boxes

TiefVision: unsupervised image similarity



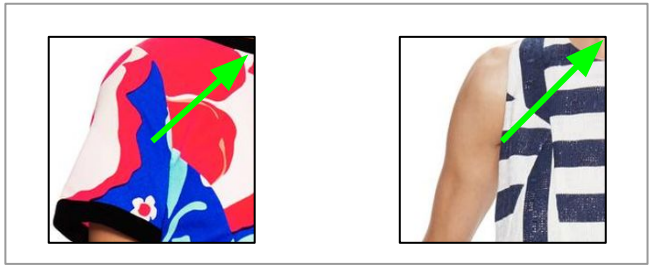
Unsupervised Image Similarity

- Get a new image by cropping the bounding box that comes from image location.
- Get the normalized encoding for each coordinate in the new image.
- The similarity is based on the average of the angle between each encoding.
- As encodings are normalized, the dot product (cosine) is used as similarity metric instead of the angle (small angles or big cosines imply high similarity)

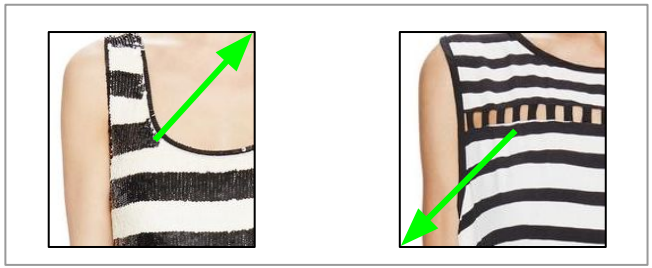
TiefVision: supervised image similarity using Deep Rank

The unsupervised model doesn't always do a good job detecting similarities:

- There will be **small encoding angles** coming from very **different crops**:



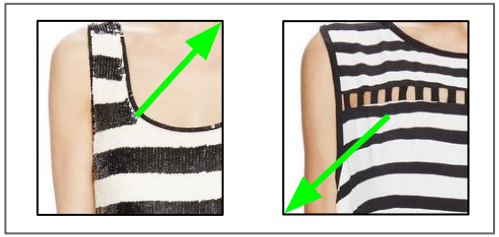
- There will be **big encoding angles** coming from very **similar crops**:



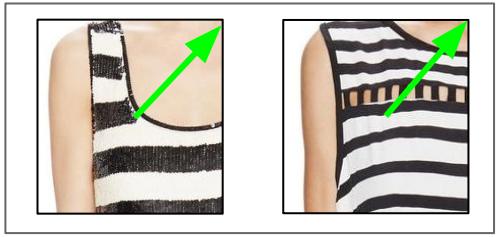
TiefVision: supervised image similarity using Deep Rank

The goal is to transform the encodings in such a way, the angles of similar crops are small and angles of different crops are big:

- Original encodings:



- Newly generated encodings coming from the output of a neural network trained in a supervised way:



TiefVision: supervised image similarity using Deep Rank

H+ (similar to H)	**H (reference image)**	H- (different from H)
		

The dataset is composed out of triplets:

- H : reference image (any image can act as reference).
- H+: an image similar to the reference image H.
- H-: an image different from the reference image H.

TiefVision: supervised image similarity using Deep Rank

We want similar crops to have smaller angles than the dissimilar crops

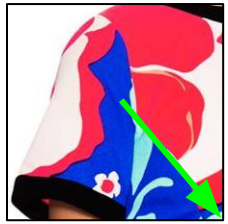
angle(*NN(*  **, *NN(***  **)) < angle(*NN(***  **, *NN(***  **))**

...and make sure the encodings angle differences are significant enough

angle(*NN(*  **, *NN(***  **)) + *margin* < angle(*NN(***  **, *NN(***  **))**

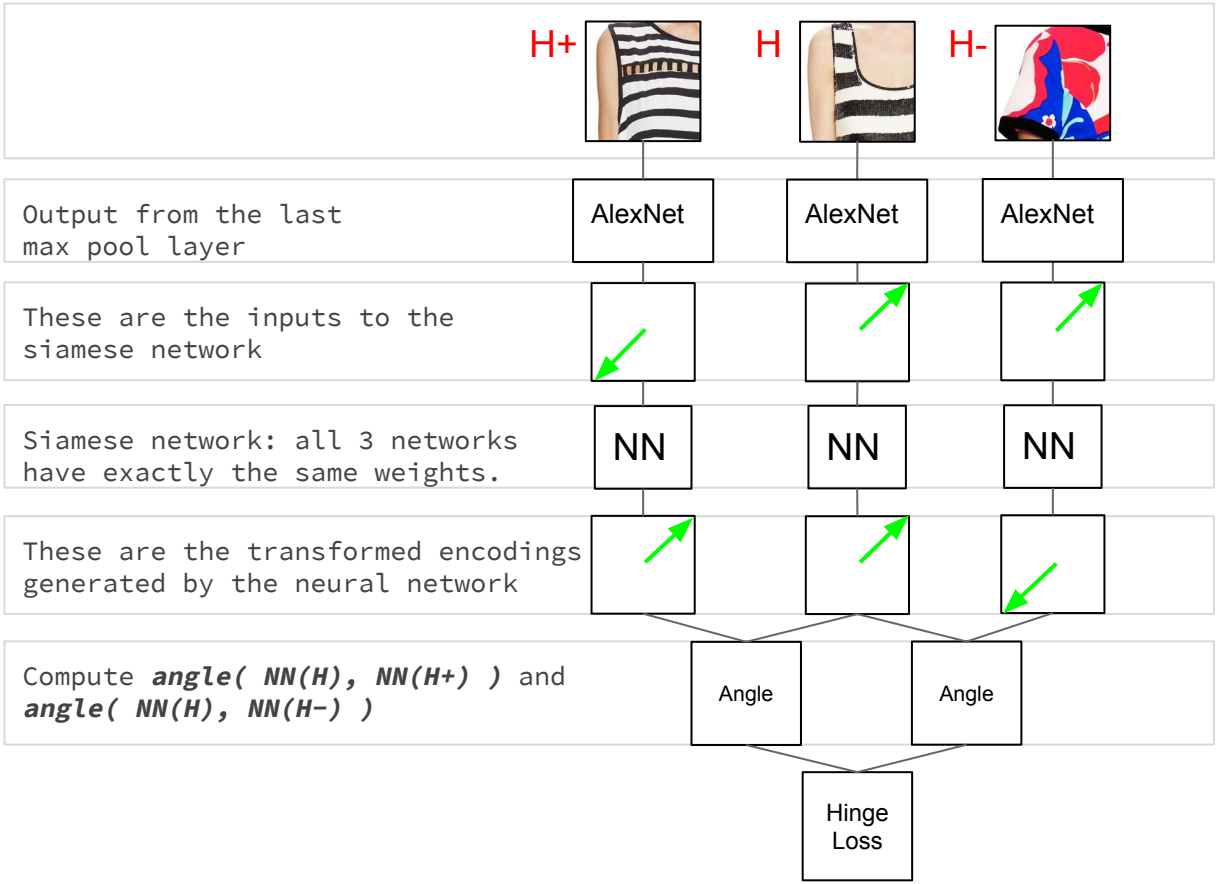
To train a network to fulfill the inequality we can use the **Hinge loss**:

max{0, margin + angle(NN( ***, NN(***  ***)) - angle(NN(***  ***, NN(***  ***)) }***

H+ (similar to H)	H (reference image)	H- (different from H)
		

Zero error with 90 degrees margin

TiefVision: supervised image similarity using Deep Rank



max(0, margin + angle(NN(H), NN(H+)) − angle(NN(H), NN(H−))

TiefVision

Any Question?

TiefVision: Papers, Articles and other links

- OverFeat: http://arxiv.org/pdf/1312.6229v4.pdf
- Deep Rank: http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42945.pdf
- Unsupervised (and also supervised) image similarity: http://research.larc.smu.edu.sg/mlg/papers/MM14-fp336-hoi.pdf
- How to convert fully connected layers into equivalent convolutional ones: http://tech.gilt.com/deep/learning/2016/05/18/fully-connected-to-convolutional-conversion
- Alexnet: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
- TiefVision: https://github.com/paucarre/tiefvision