CentraleSupélec

Where can I find this shirt?

Fatma Moalla & Mohamed Karroumi & Niraj Srinivas & François Le Roux

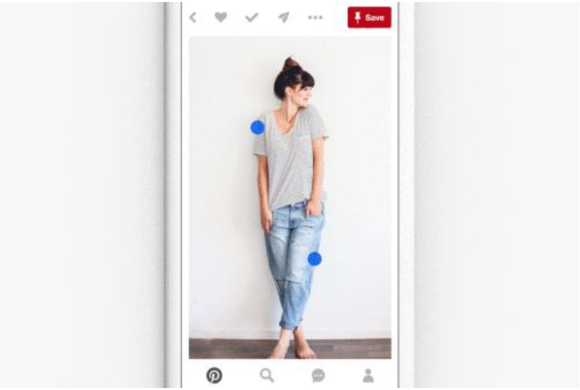
Visual Computing Final Project

-

Visual Search

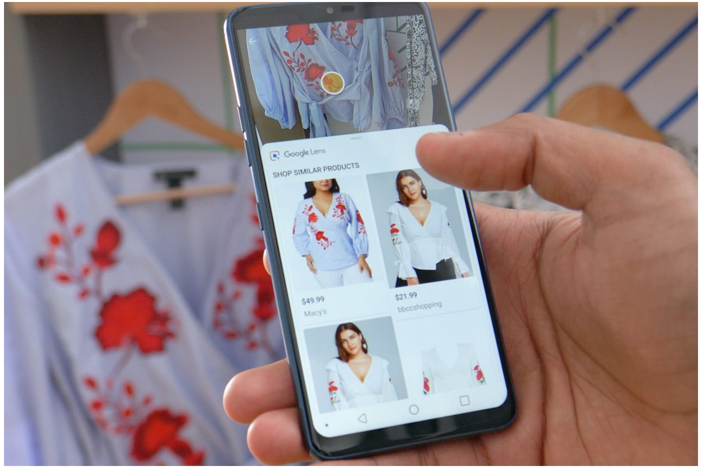
-

05/03/2020



Project Motivation

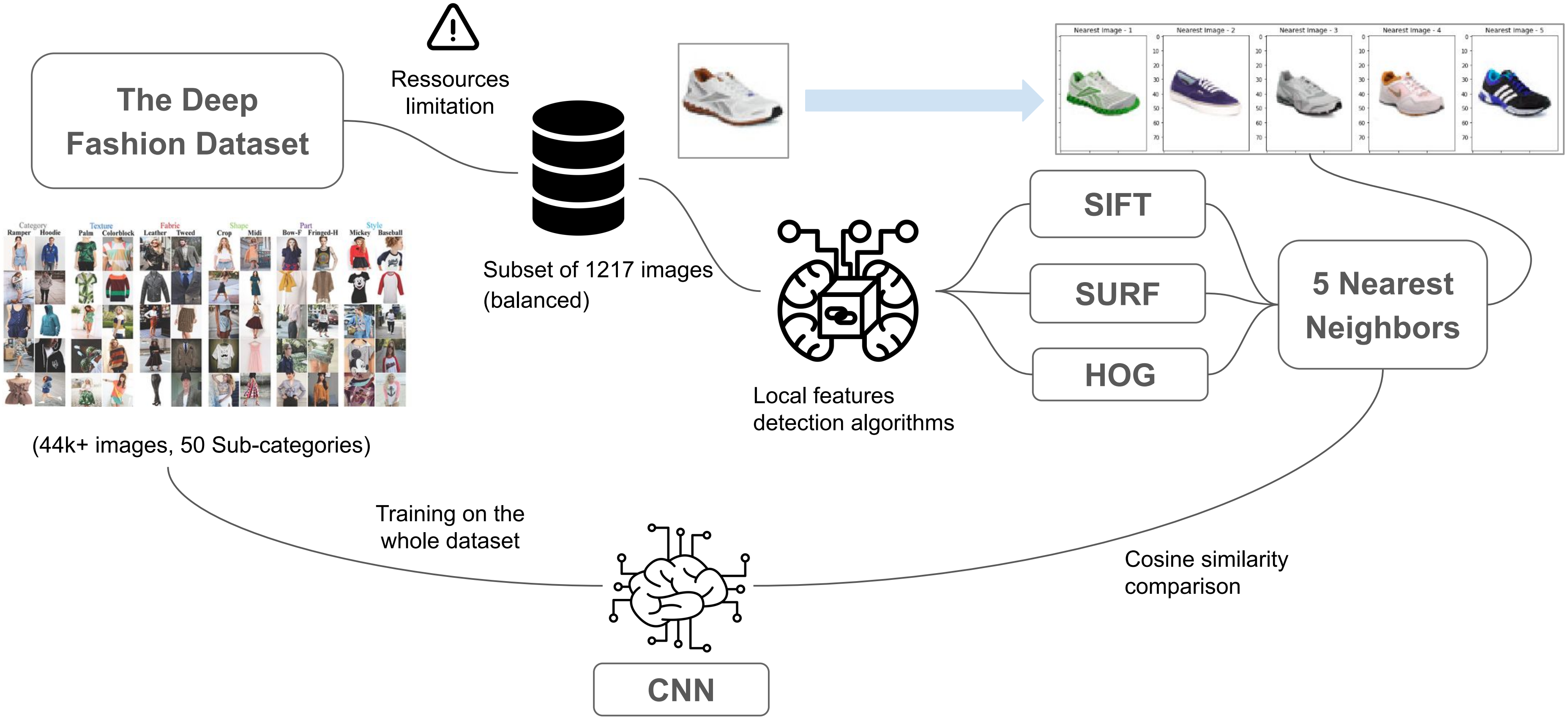
- Year 2000
- Jennifer Lopez and Versace
- “Jennifer Lopez’s green dress”, the most popular search query
- The demand for more than just text
- Birth of Google Image Search
- Fastrack to 2017, Google Lens launched



Where can I find this shirt?

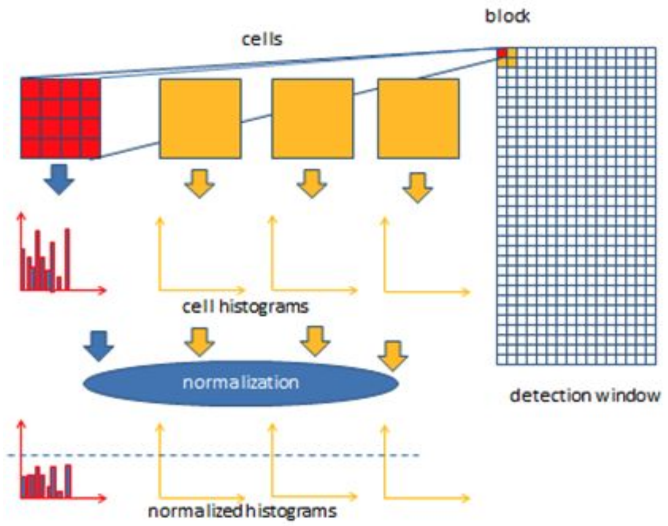
- Growing Mobile and E-Commerce segments in Fashion Industry
- Multiple choices and offers
- But Personalisation = Survival
- Styles, colors, textures and types of dresses contain more information on user’s preference than just a few words
- Recommend products based on users’ preferences and history of product search.

Model Pipeline



Histogram Of Gradients (HOG)

Main objective: Converts the image to a feature vector → Important for object recognition

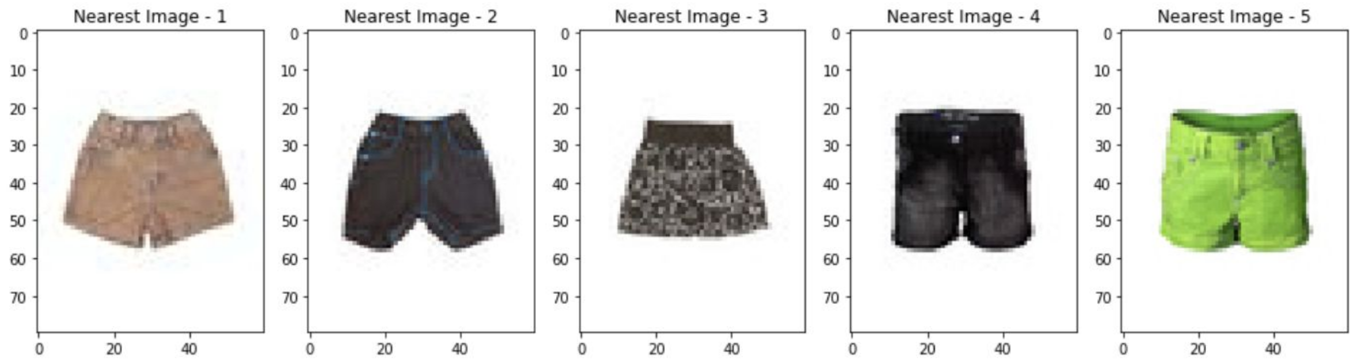


Keys steps

1. **Image preprocessing: patches and cells creation**
2. **Gradient (Sobel):** magnitude + direction
3. **Histogram:** assign a bin (angle) to each cell
4. **16 x 16 Block Normalization**
5. **Concatenate blocks:** a single feature vector (block_size x number of positions)

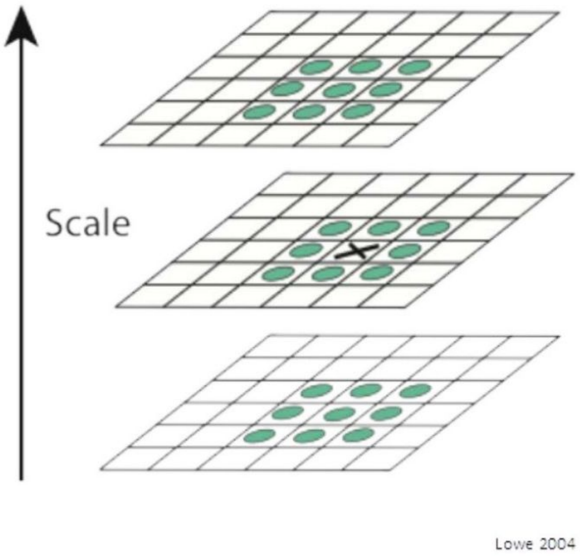
HOG results on the Deep Fashion dataset

- Cosine similarity = 96%
- Visual Matching
- Detects texture information
- Computationally not intensive



Scale-Invariant Feature Transform (SIFT)

Main objective: building scale invariant image descriptors



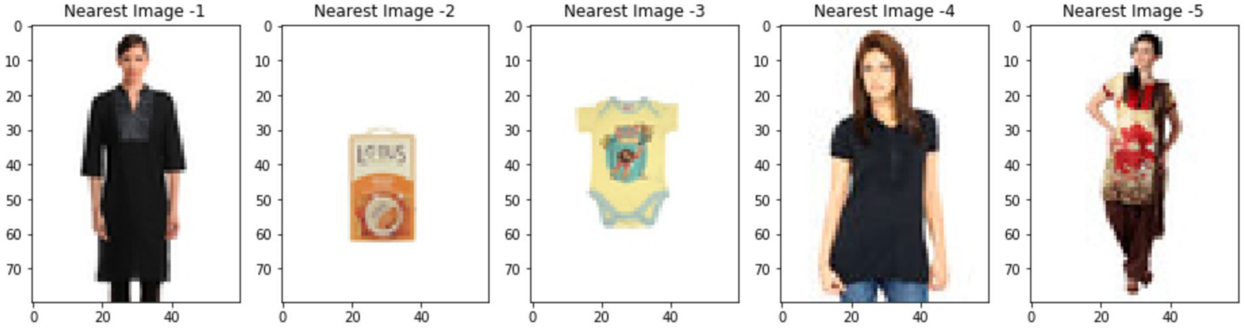
Keys steps

1. **Key point detection:** LoG
2. **Feature point localization:** Difference of Gaussian (DoG)
3. **Orientation assignment:** uses Histogram of Orientation
4. **Feature descriptor generation**

SIFT results on the Deep Fashion dataset

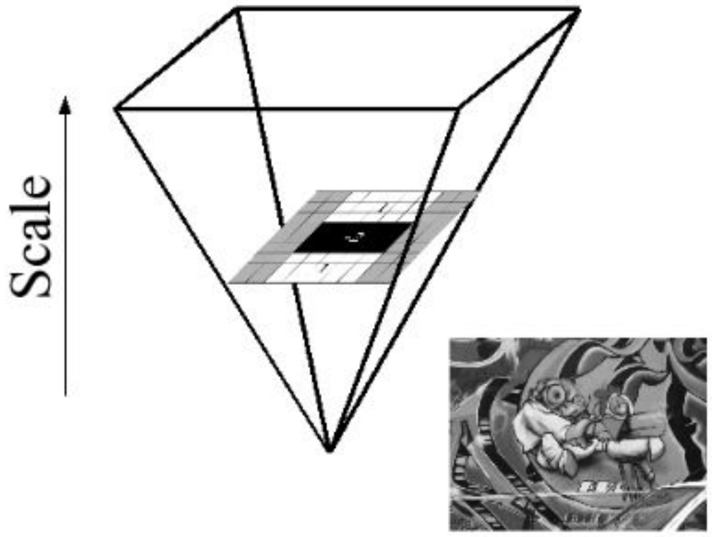
- Cosine similarity = 91.2% ! ✓
- Visual Matching

→ **Biased dataset or too small images (80 x 80)**
→ **Wrong metric**



Speeded Up Robust Features (SURF)

Inspired by SIFT, but designed to be faster



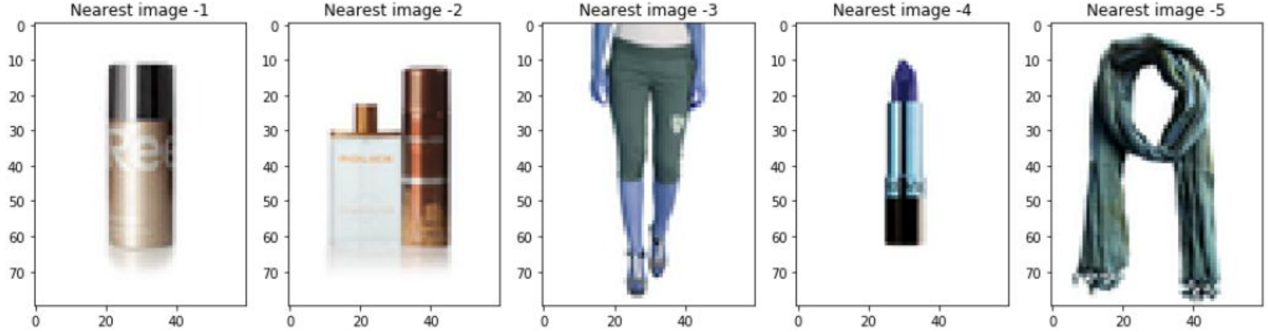
Keys steps

1. **Key point detection:** Integral image, box filters to approximate Hessian determinant
2. **Selection:** box filters of different sizes for scale, non-max suppression
3. **Orientation assignment:** Haar wavelet response on sliding windows of size $\pi/3$
4. **Feature descriptor generation:** Haar wavelet responses weighted by Gaussian

SURF results on the Deep Fashion dataset

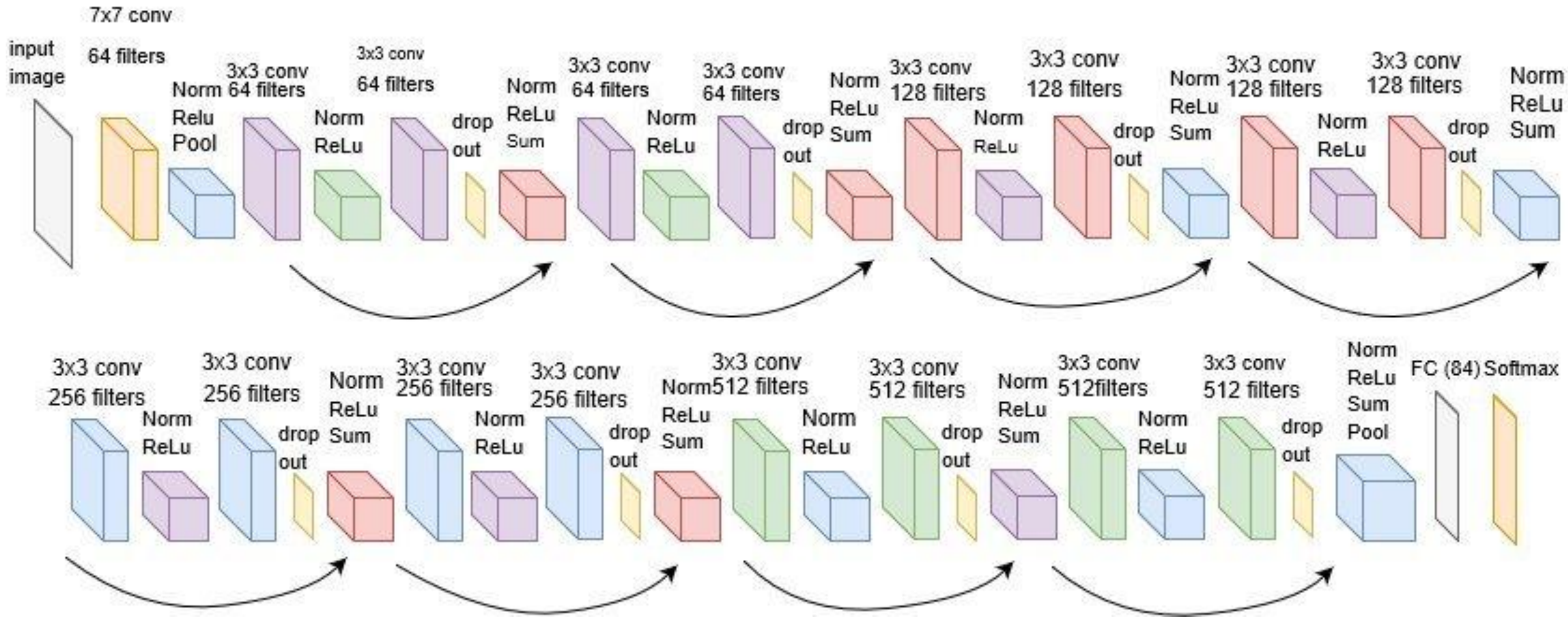
- Cosine similarity = 94.6% ! ✓
- Visual Matching

→ Biased dataset or too small images (80 x 80)
→ Not optimal matching method



Convolutional Neural Networks

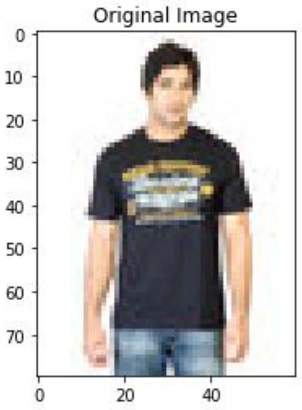
- Pretrained model: ResNet_18



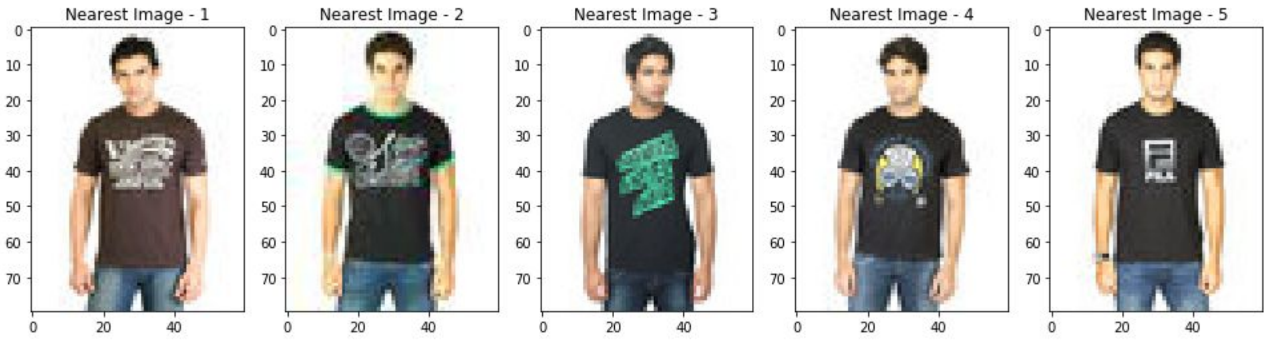
Convolutional Neural Networks: Approach

- We take a pretrained model: ResNet_18
- We fine tune the weights of the model to perform Classification on Deep Fashion Dataset.
- For each image extract the second-last layer as a Feature map of the image.
- Use LSHashing on feature maps to find the nearest neighbor.

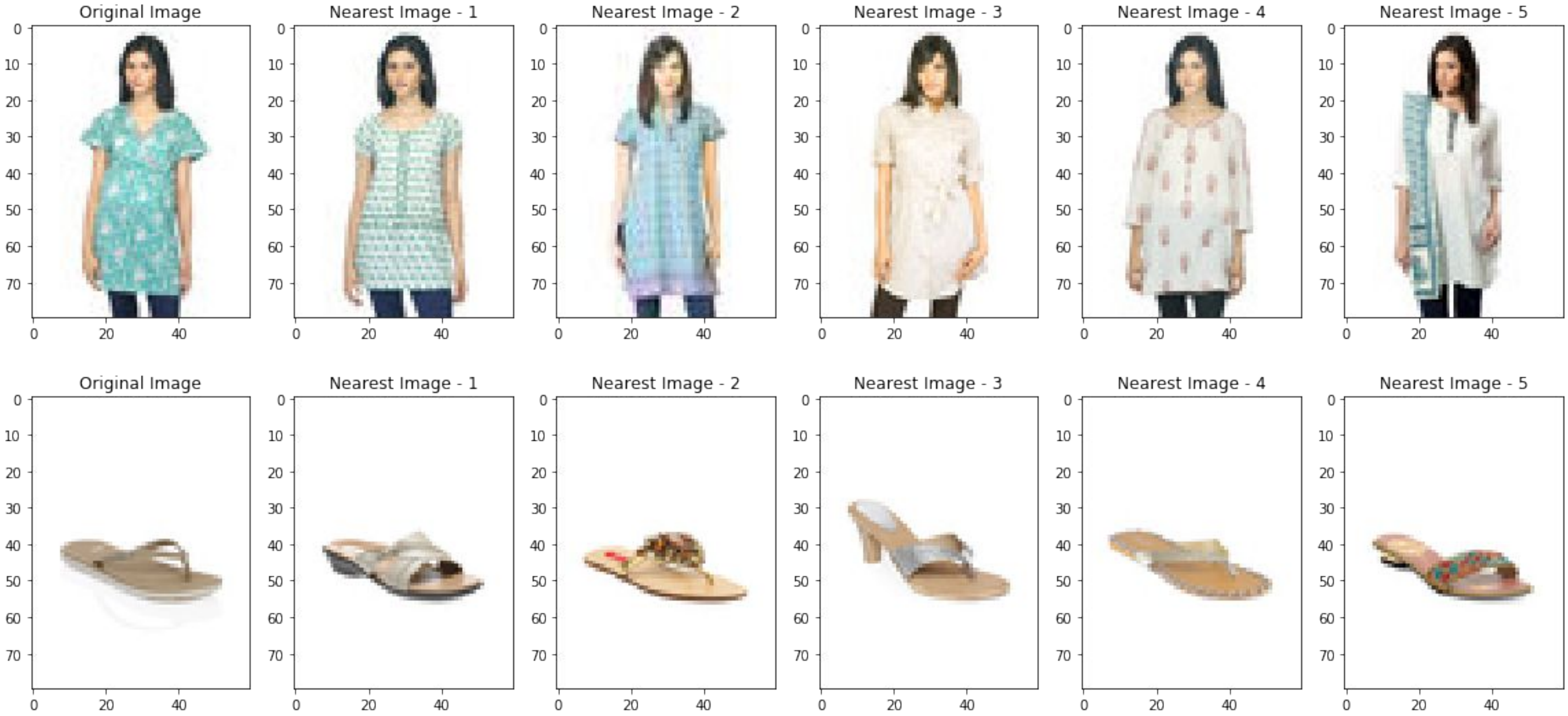
CNN results on the Deep Fashion dataset



- Cosine similarity = 98.11% ✓
- Visual Matching
- Detects texture information
- Computationally intensive



More Results



Final results comparison

Approach	HOG	SIFT	SURF	CNN
Average Cosine Similarity	96.3%	93.4%	94.6%	**98.11%**
Visual Matching	Good	Very Bad	Medium	**Perfect**

Final discussion and Conclusion

- **Challenges**: scalability to the whole dataset, accuracy of features for small images
- **Take home messages**: HOG works about as well as CNN for this type of application, SIFT and SURF have some difficulties
- **Next steps**: Try on harder images (non uniform background, more varied situations depicted), try on a larger dataset

Thank you for your attention