Modern CNNs & transfer learning

Alexander Ecker

Face recognition using CNNs

Face Recognition

Database of *K* persons

Get an input image

Output the ID of the person in the image

Face Recognition

One challenge: few-shot learning

- Number of persons is large and the number of images per person small
- We need to recognize a person given just one or a few examples

Simply training a classifier is not a good solution

- It won’t be good enough to recognize a person from few images
- Adding a new person will require retraining the network

Face Recognition

One solution: learn a similarity function instead of a classifier

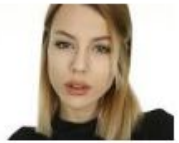
At inference time:

- Given a new image, compute how similar it is to images in the database
- If similarity is greater than threshold
  - person is present in the database
  - output id of closest sample (or *k* nearest neighbors)

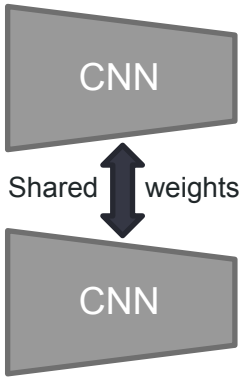
Can use *Siamese networks* with *contrastive learning objectives* to solve this issue

Face Recognition: Siamese networks

$x^{(1)}$



$x^{(2)}$



$f(x^{(1}$

$f(x^{(2}$

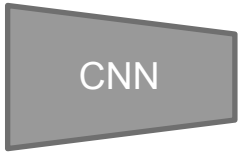
Distance function

(opposite of similarity)

$$d(x^{(1)}, x^{(2)}) = \|f(x^{(1)}) - f(x^{(2)})\|_2^2$$

Goal of learning

$x^{(1)}$



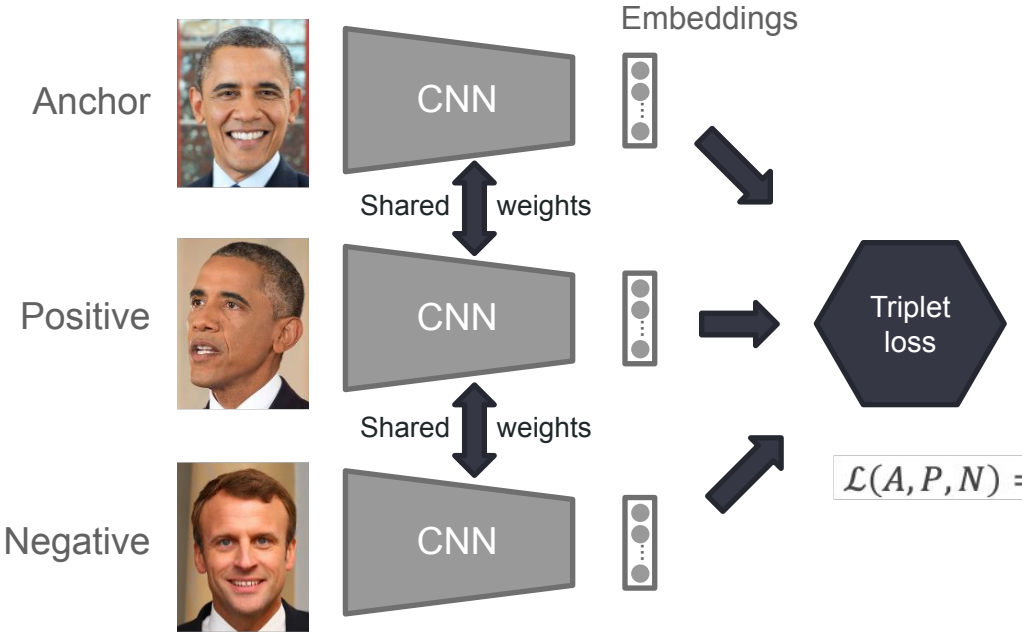
$f(x^{(1}$

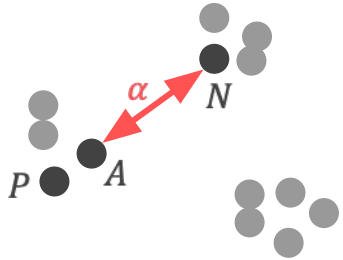
Parameters of CNN define an embedding $f(x^{(1)})$.

Learn parameters such that:

- If $x^{(i)}, x^{(j)}$ are the same person, $\|f(x^{(i)}) - f(x^{(j)})\|_2^2$ is small
- If $x^{(i)}, x^{(j)}$ are different persons, $\|f(x^{(i)}) - f(x^{(j)})\|_2^2$ is large

Learning objective: triplet loss





Minimize!

$$\mathcal{L}(A,P,N) = \max(\|f(A) - f(P)\|_2^2 - \|f(A) - f(N)\|_2^2 + \alpha, 0)$$

small large

Exercise

Visualize embeddings learned by
the face recognition network VGG-face

Does it separate the pictures of different people?

Ethics of deep learning

Example: face recognition and dataset bias

Face recognition is super useful…

…but can also be problematic

Should we use (or even develop) facial recognition technology?

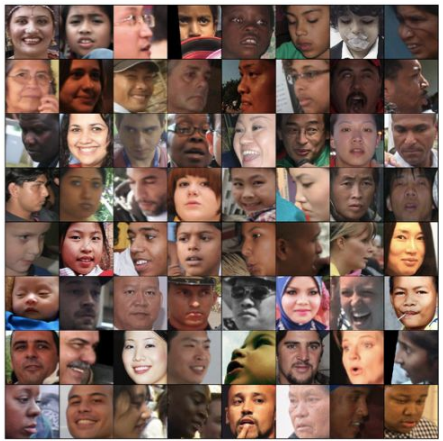
That’s an interesting question I encourage everybody to think about

Focus here: more subtle technical issues with face recognition
(important, since it’s already being deployed)

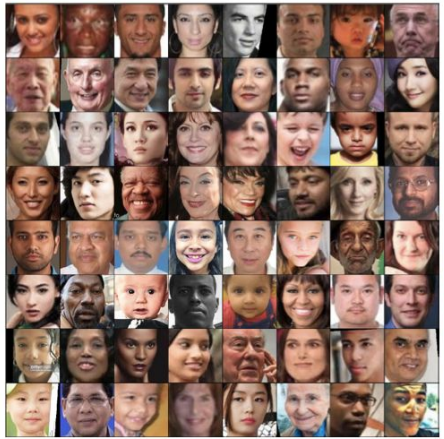
Notice anything?

Karkkainen & Joo:
“FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation.”
IEEE/CVF Winter Conference on Applications of Computer Vision 2021

Many face datasets are heavily biased



(a) FairFace



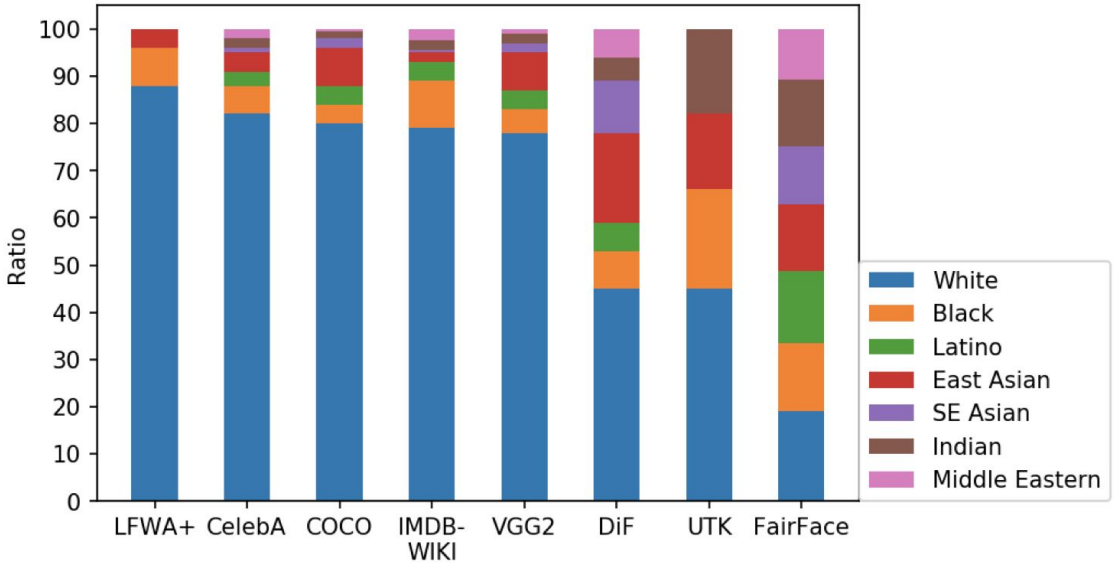
(b) UTKFace

(c) LFWA+

(d) CelebA

Karkkainen & Joo: 2021

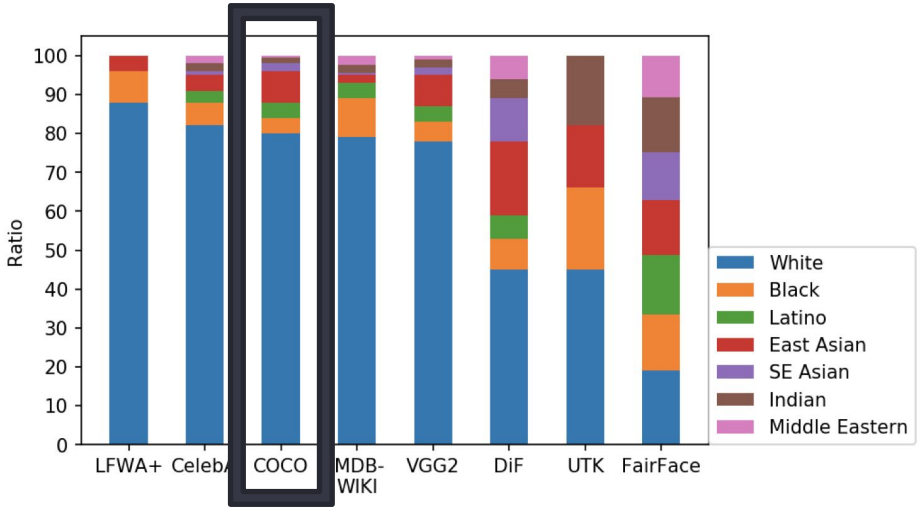
Many (face) datasets are heavily biased



Karkkainen & Joo: 2021

Not just a problem of face recognition

Other datasets (e.g. COCO, ImageNet) have similar biases



What does it mean?

People who are not white or from western cultures
are not equally represented in the data

- Algorithms are more likely to make mistakes for them

- We should be very conscious about the datasets we use
  and the consequences it might have down the road

Exercise

Explore how dataset bias affects the learned representations.

How much space do black vs. white people “occupy” in the learned representation?

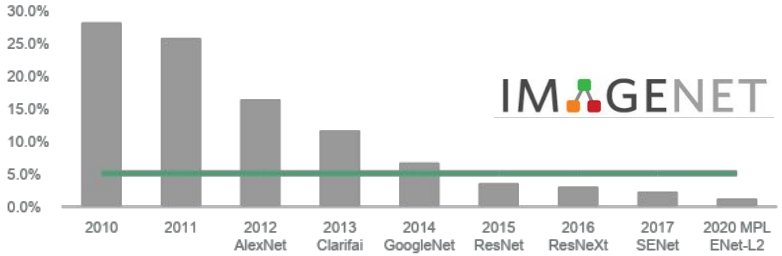
How might that affect systems that employ transfer learning for recognizing people?

Summary and outlook

Beyond convnets and supervised learning on ImageNet

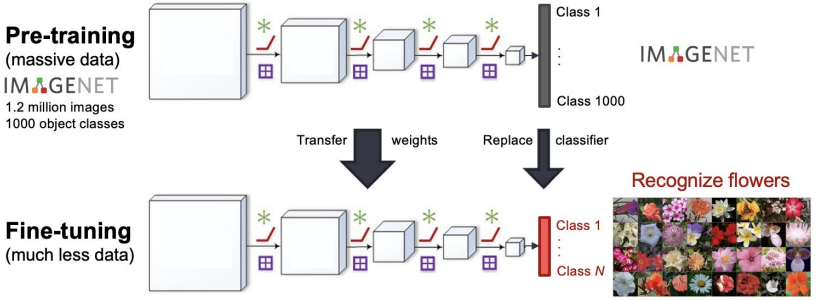
Summary

ImageNet and large-scale CNN architectures



Face recognition

Transfer learning



Summary

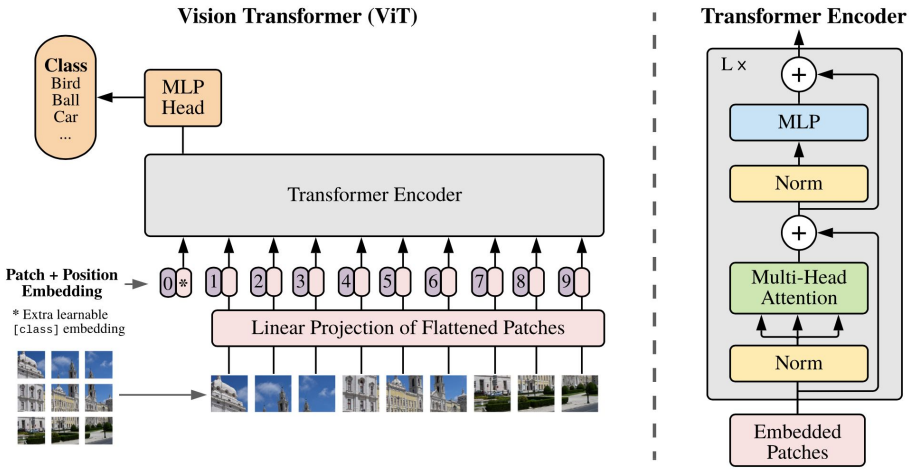
Large-scale CNNs and transfer learning based on ImageNet pre-training have been instrumental for the deep learning revolution after 2012.

What's next?

Beyond CNNs: Vision Transformers

Transformers have completely taken over natural language processing a few years ago

They have also become increasingly popular in image recognition and now perform on par with CNNs

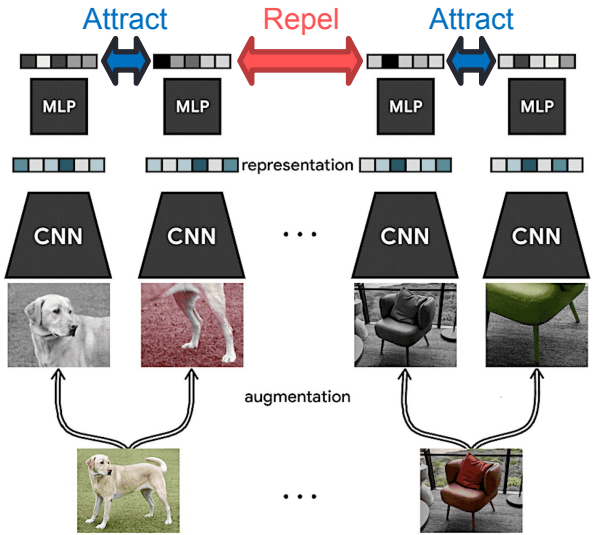


Vision Transfermer (ViT)
Dosovitskiy et al. 2020

Beyond supervised pre-training on ImageNet

Self-supervised learning has made tremendous progress recently

Now outperforms supervised pre-training on ImageNet on some transfer tasks such as object detection



SimCLR. Chen et al., ICML 2020
https://ai.googleblog.com/2020/04/advancing-self-supervised-and-semi.html

Conclusion

Large-scale CNNs and transfer learning based on ImageNet pre-training have been instrumental for the deep learning revolution after 2012

It is still the standard approach, but the paradigm may shift again in the next couple of years