Machine Learning Project - Hand-To-Age (H_2A)

Luis Chaves

MSc Health Data Analytics & Machine Learning

10/03/2020

Project description

Data

Data source: Radiology Society of North America(RSNA) and Radiology Informatics Committee (RIC). Available in Kaggle. Images gathered by several

Dataset: 12,621 images of individuals aged between 1 month and 19 years (228 months) old. Gender and age available for all fo them.

Context: Images gathered for the Pediatric Bone Age ML Challenge.

Aim(s) of my study

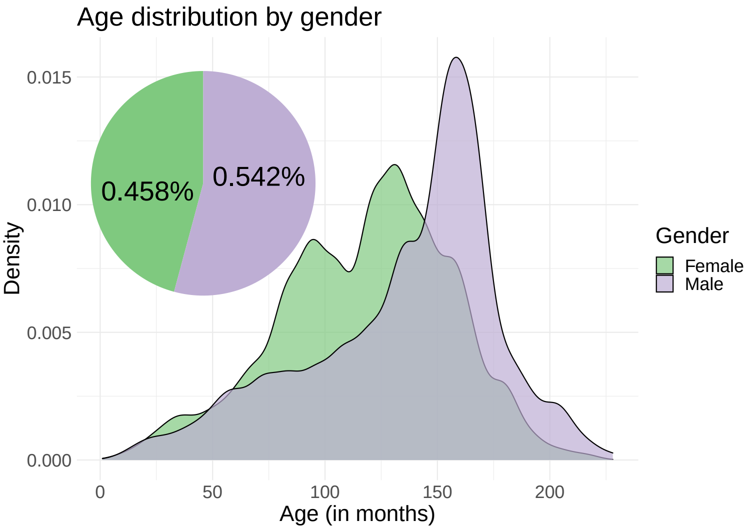
Supervised question #1: *How close can we estimate age from images only?*

Unsupervised question#1: *Can clustering algorithm accurately group together individuals by gender*

Unsupervised question#2: *Unsupervised learning for out-of-the-box applications in computer vision*

(Appendix) Supervised question #2: *Can gender be derived from the image?*

Population statistics



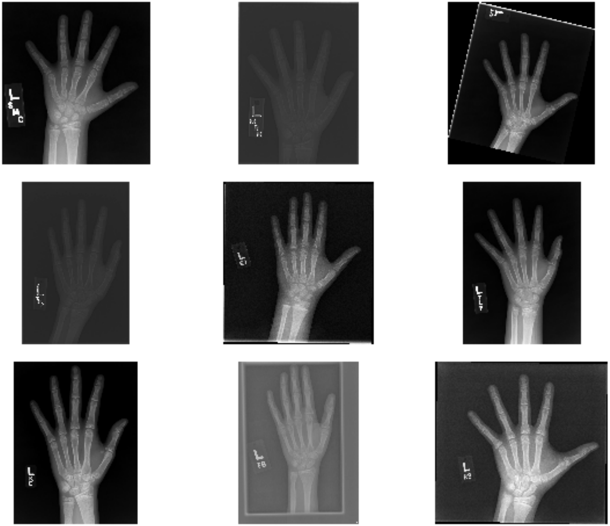
The images

X-ray images of each individuals' hand (one or two - information not available)

- Difficulties:
  - Varying resolution
  - Varying contrast
  - Varying exposure
  - Some scanned and some digital images
- Advantages:
  - Standardised medical images

Let's have a look at some pictures!

Raw Images



Raw Images

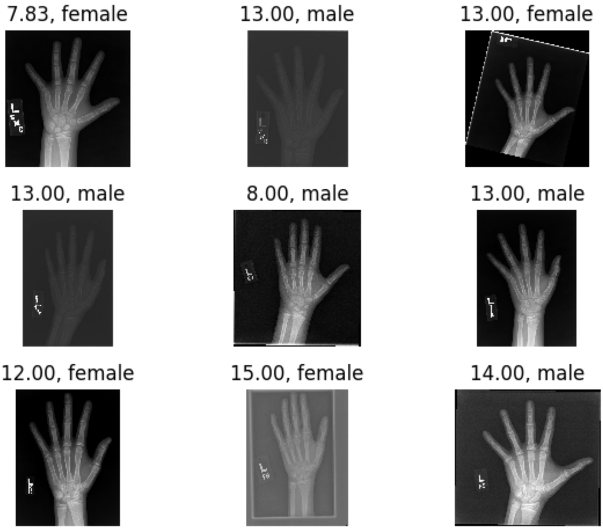

7.83, female
13.00, male
13.00, female
13.00, male
8.00, male
13.00, male
12.00, female
15.00, female
14.00, male

Image processing & Feature engineering

Data processing

Data split

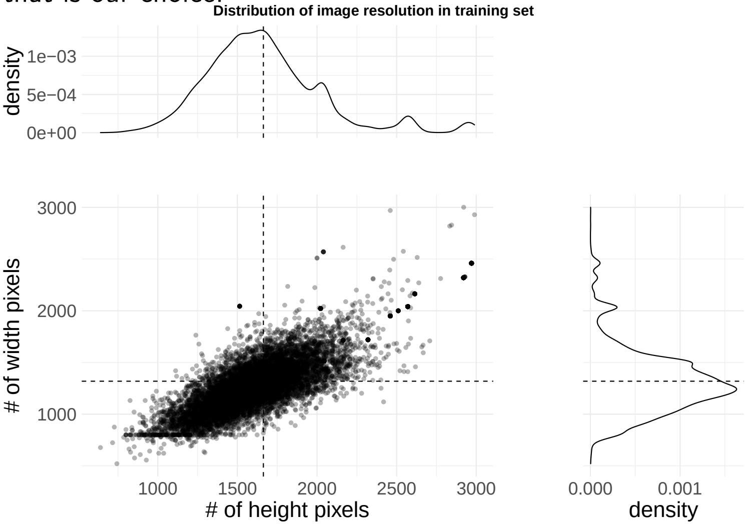
10,000 Images for training, 2611 for testing/validation, no cross-validation because of computational cost and large amount of training data.

Image processing overview

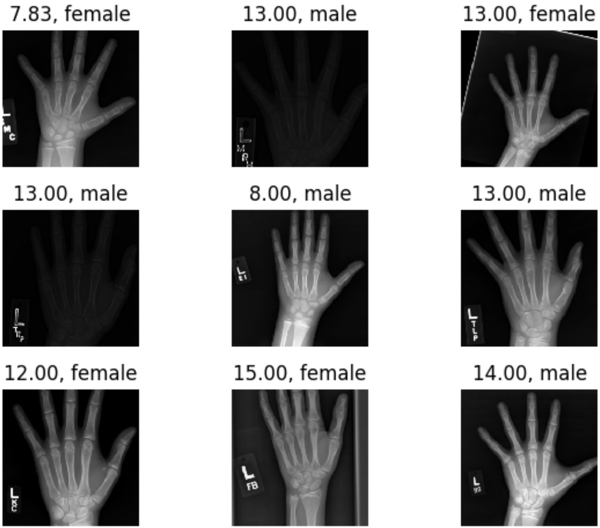
- Rescaled and Center Cropped Images
- Centering and scaling features (pixel values)
- Contrast adjustment

Large disparity in image resolution

Default for many network architectures is 224x224 so that is our choice!



Rescaling and cropping

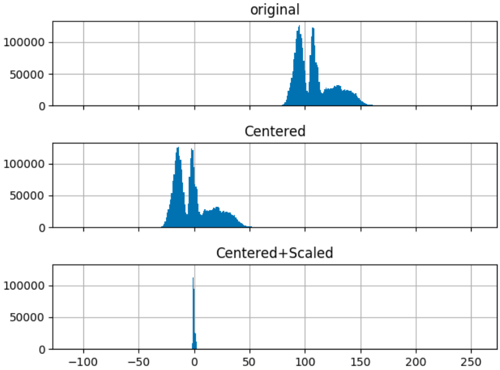


Centering and scaling

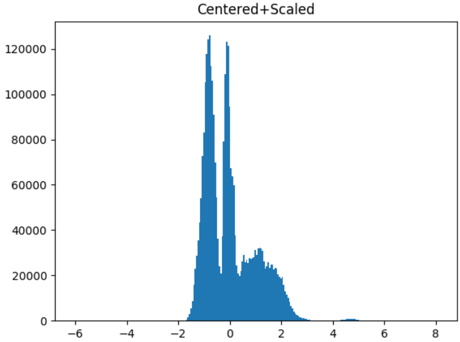
It is widely recommended to center and scale the inputs of a neural networks as this will speed up training by gradient descent (see notes from CS231n Stanford).
Several forms of normalisation: Instance Normalisation(IN) and Batch Normalisation(BN):

- **IN: Center and scale each image w.r.t. its mean and standard deviation.**
- **BN: Center and scale each image w.r.t. the whole batch of images mean and standard deviation**

Centering and scaling



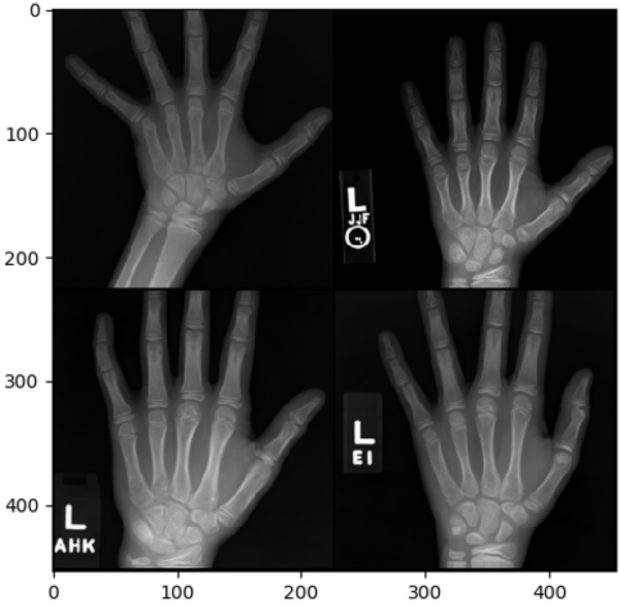
Centering and scaling



Until further noticed all results from neural networks come from images centered with the mean and scaled with the standard deviation.

Note! where images have low exposure, most normalised pixel value will become negative and clipped to 0 when plotted!

For reference, images look like this



Deep Learning for Computer Vision

Network of choice: ResNet

Residual networks were introduced by reserachers at Microsoft in 2015 (see He et al. (n.d.)). They represented a breaktrough at the time because they allow to train deeper network with an added trick which added no cost to the training computationally. ResNets are made of blocks such as the one below:

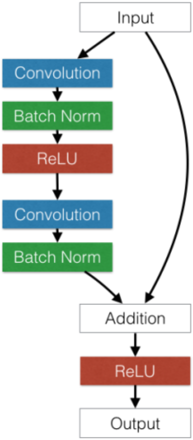
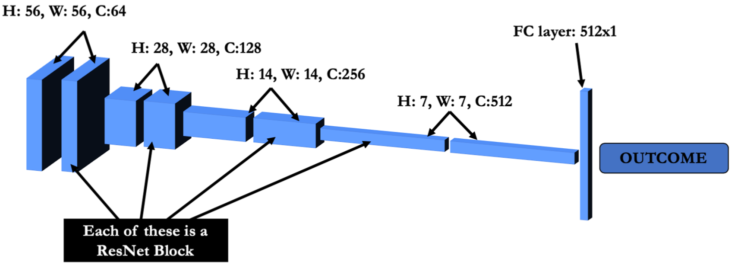


Figure 1. A RestNet basic block

Network of choice: ResNet

Residual networks were introduced by reserachers at Microsoft in 2015 (see He et al. (n.d.)). They represented a breaktrough at the time because they allow to train deeper network with an added trick which added no cost to the training computationally.
Looking at the overall architecture it looks like this:



Hyperparameter tuning

- Learning rate α
- Optimizer: **SGD**, Adam
- Learning rate scheduler: StepLR, Exponential LR, ReduceONPlateauLR, **CyclicLR**, **OneCycle Policy**
- Image normalisation: batch vs instance
- Networks' depth (# of layers): 18, 34 and 50 layer-deep ResNets
- Regularisation: L2-regularisation

Hyperparameter tuning

- Learning rate α
- Optimizer: **SGD**, Adam, Adagrad, AdaDelta, AdamW, SGD with momentum...
- Learning rate scheduler: StepLR, Exponential LR, ReduceONPlateauLR, **CyclicLR**, **OneCycle Policy**

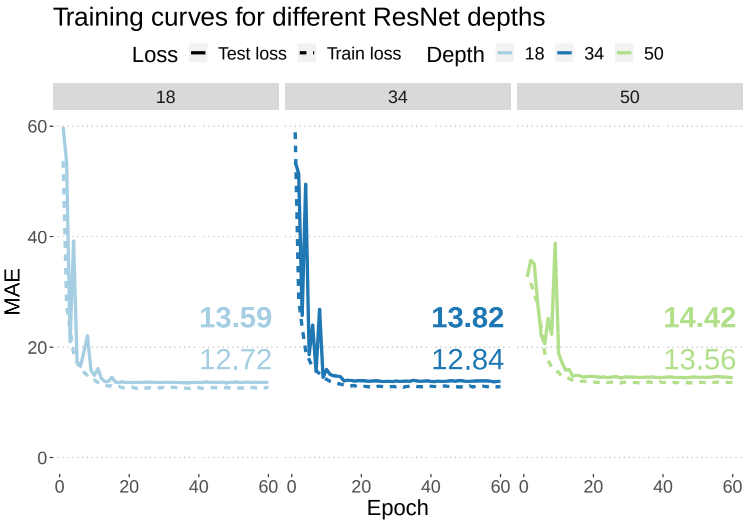
"This (α) is often the single most important hyperparameter and one should always make sure that it has been tuned" (see Neural Networks: Tricks of the Trade - 2012)

Exponentially growing possibilites

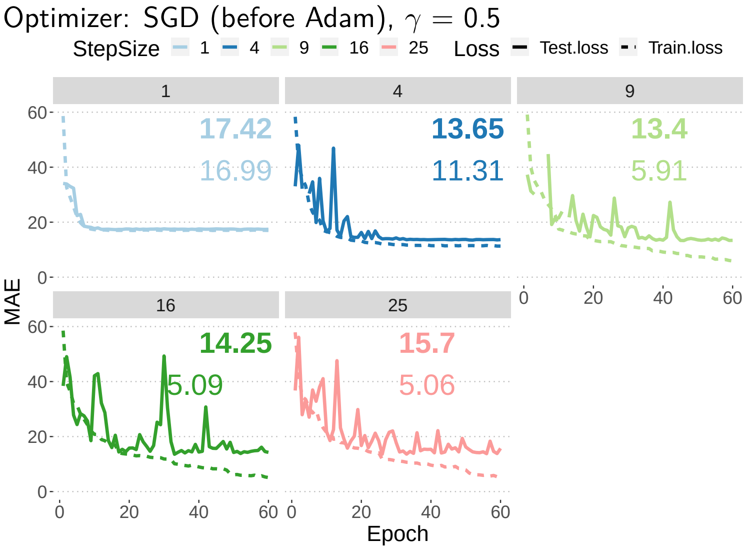
The ‘Caviar’ approach: training many many models with different hyperparameters, seeing which one does best. I tried:

- Varying ResNet Depth: 18, 34, 50 layers.
  - IN, $\alpha = 1$,ExponentialLR schedule, $\gamma = 0.7$, ADAM optimizer, no weight decay
- Changing LR schedule: ReduceOnPlateau, StepLR with different γ (0.1 and 0.5) and stepsize (1,4,9,16,25).
- **Moving into smarter ideas:** CyclicLR, LR Range finder and One cycle policy.

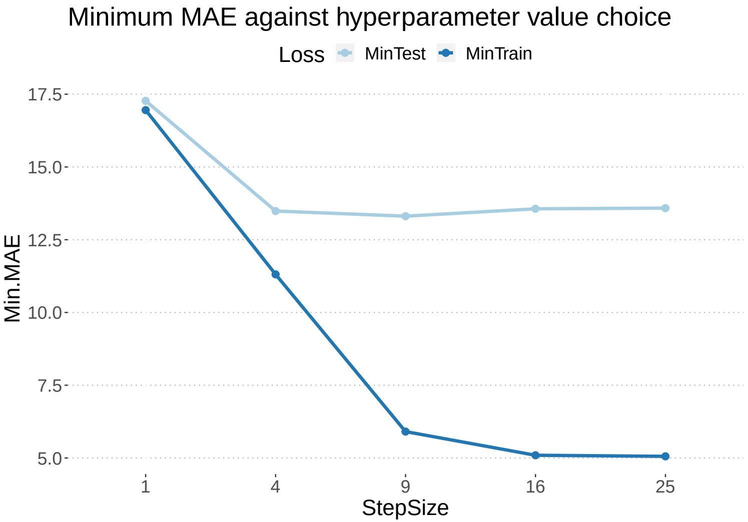
Deeper network does not lead to better results



Varying step-size leads to different results



Exploring the bias-variance tradoff

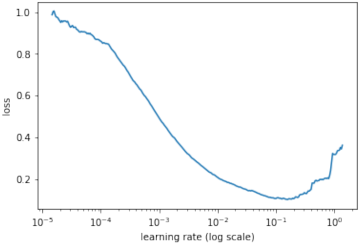


Challenging conventional wisdom

Cyclic Learning Rates and One-Cycle Policy

Based on work by Leslie Smith (Smith (2015), Smith and Topin (2017) and Smith (2018))

LR range test:
During one (or a few) epochs, we train a given network with increasing α for each training iteration (i.e for each batch).

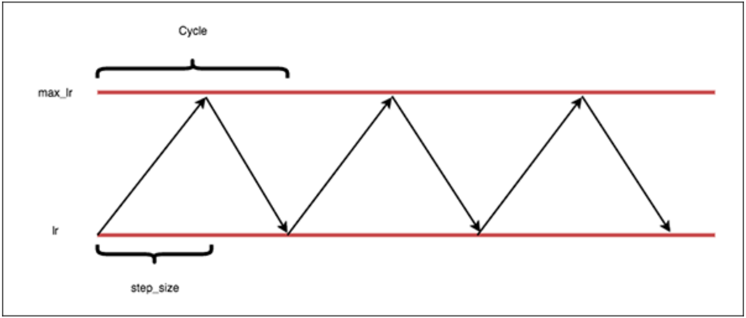


Challenging conventional wisdom

Cyclic Learning Rates and One-Cycle Policy

Based on work by Leslie Smith (Smith (2015), Smith and Topin (2017) and Smith (2018))

CLR:
Learning rate oscillates between two bounds found with the LR Range test

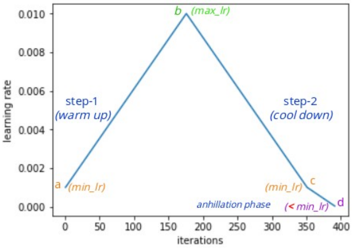


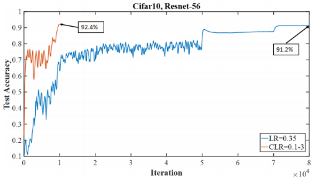
Challenging conventional wisdom

Cyclic Learning Rates and One-Cycle Policy

Based on work by Leslie Smith (Smith (2015), Smith and Topin (2017) and Smith (2018))

OneCycle policy:
Modification of the CLR method where there is only a single cycle with α decreasing further than the lower bound in the last iterations.

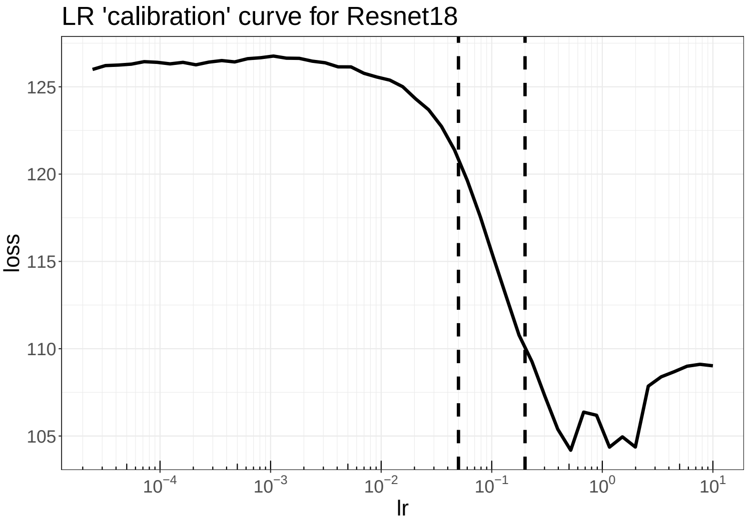




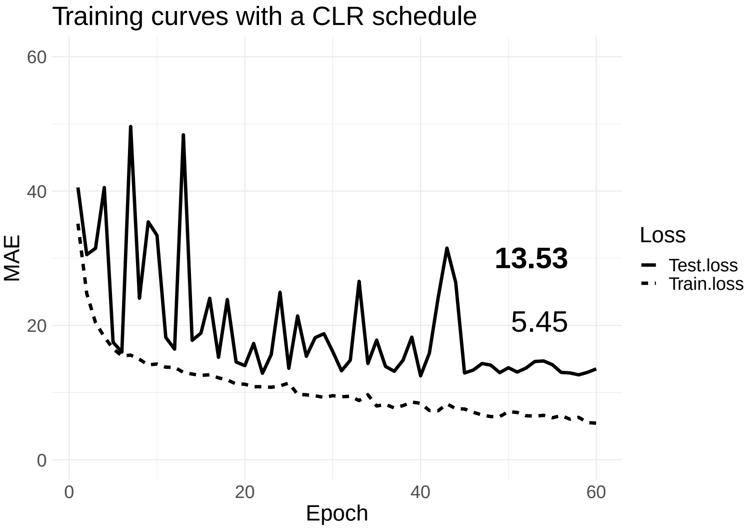
(a) Comparison of test accuracies of super-convergence example to a typical (piecewise constant) training regime.

My results with the previously mentioned methods - LR Range test

SGD, no regularisation, no momentum

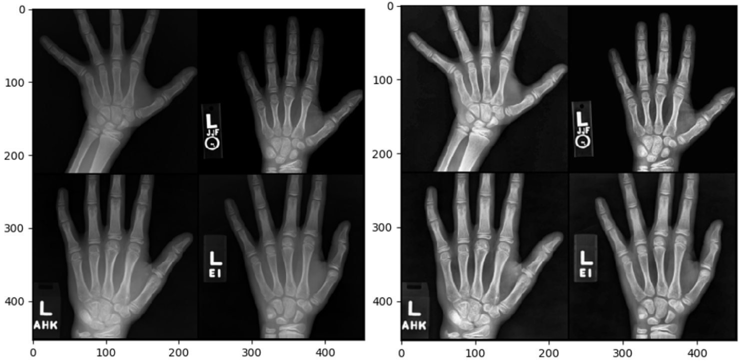


My results with the previously mentioned methods - CLR



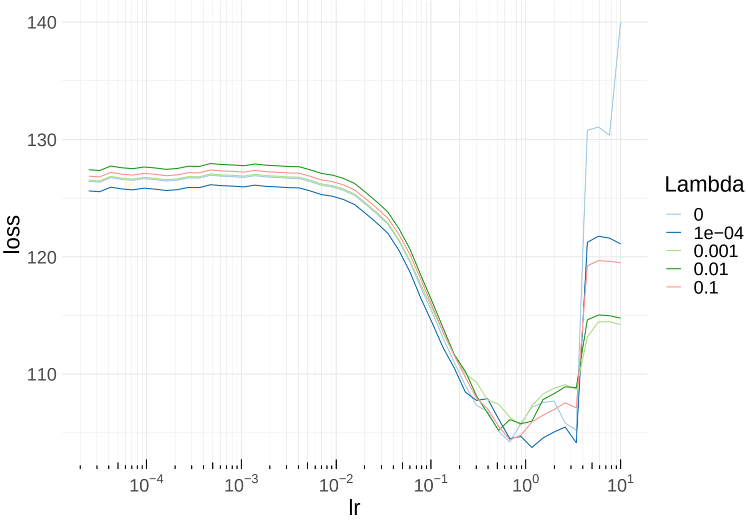
Last minute surprise!

I discovered this great image processing tool called CHALE (a kind of adaptive histogram equalisation algortihm) which makes the images sharper (more in appendix if interested)



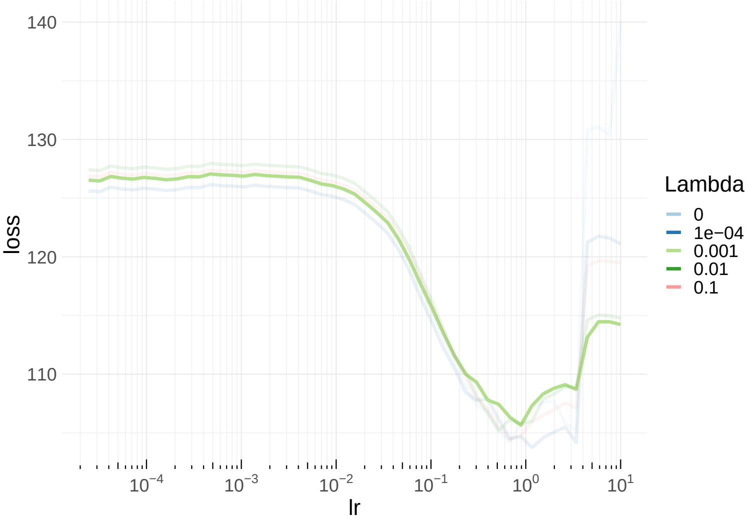
Incorporating CHALE and the OneCycle policy

Incorporating CHALE for shaper images as well as regularisation L2 norm and OneCyclePolicy



Incorporating CHALE and the OneCycle policy

Incorporating CHALE for sharper images as well as regularisation L2 norm and OneCyclePolicy



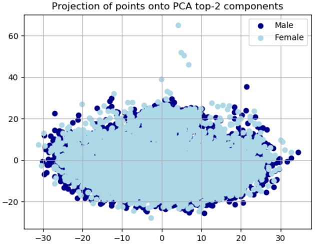
Results from 1-Cycle Policy with CHALE

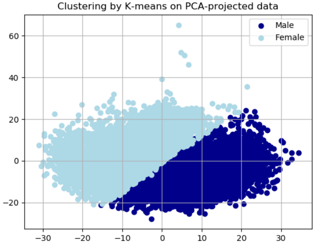
. . . A work in progress?

Unsupervised Learning

Unsupervised learning for clustering by gender

Not much difference between females and male hand bones!



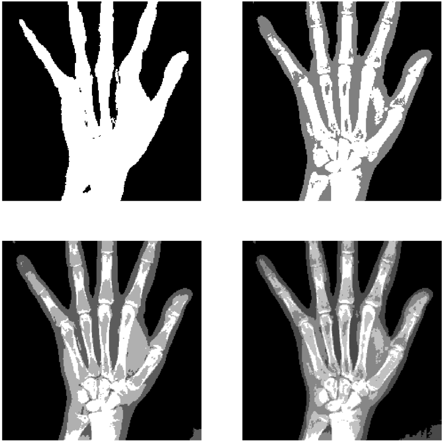


Method:

1. Decorrelate data through PCA
2. Keep components that ensure 90% explained variance
3. Perform K-Means with 2 clusters

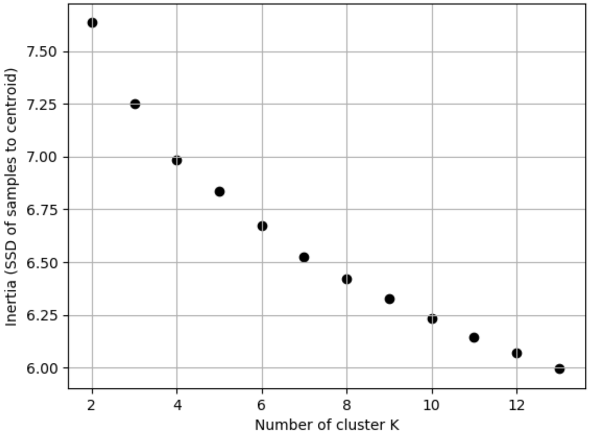
Unsupervised learning for image segmentation (proof of concept)

Using k-means to cluster region of images and using the ‘elbow’ method to assess which number of clusters is ideal.



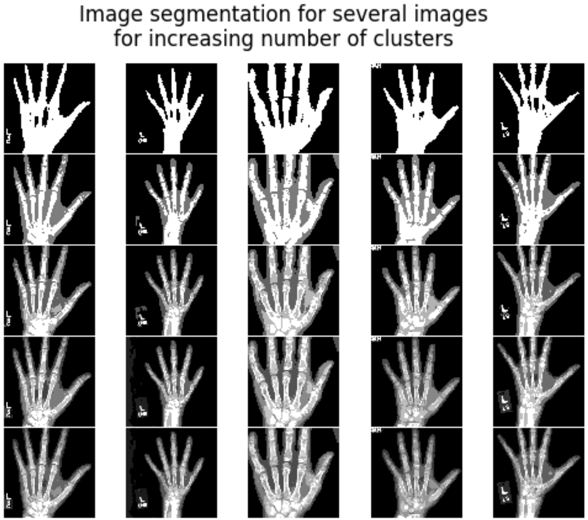
Unsupervised learning for image segmentation (proof of concept)

Using k-means to cluster region of images and using the 'elbow' method to assess which number of clusters is ideal.



Unsupervised learning for image segmentation (proof of concept)

Now testing for a few images with several number of clusters.



Discussion

Discussion

+ I achieved a MAE of 10 months with a large validation set.
+ I explored a range of optimisation algorithms and learning rate policies
+ I explored interesting applications(segmentation, compression) of unsupervised learning to medical images

− Discriminating females and males from hand ray images was not possible with deep learning nor unsupervised learning methods

Limitations and further work

Limitations and further work

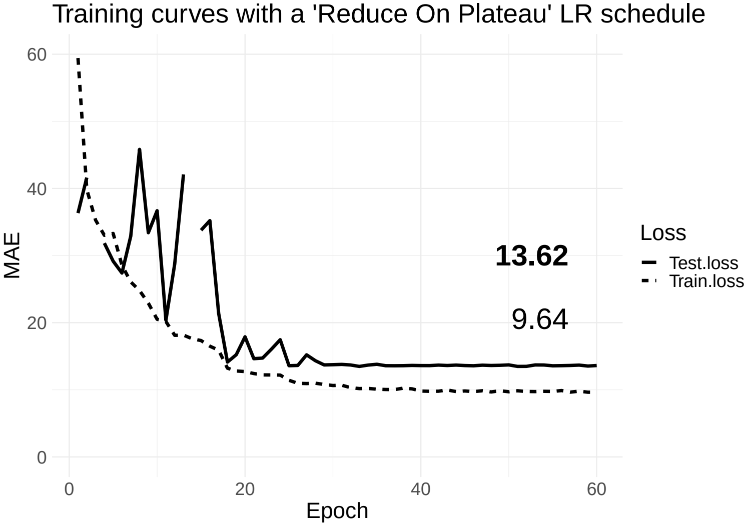
- Turning regression problem into classification by binning individuals into age groups
- Trying different neural net architectures, especially simpler ones
- Deep learning clustering methods such as VAE and t-SNE
- Incorporating sex information for age prediction
- Using images with higher resolution

Appendix

Unsupervised learning for image compression

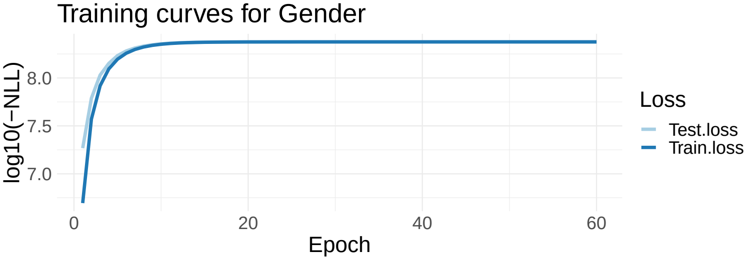
Instead of fixed learning rate changes, reduce on plateau - a slightly smarter choice

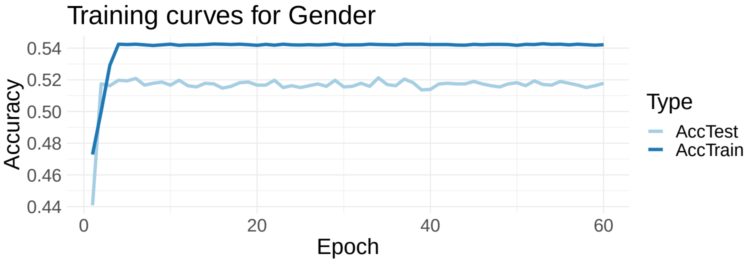
Optimizer: SGD, patience: 5, γ: 0.25



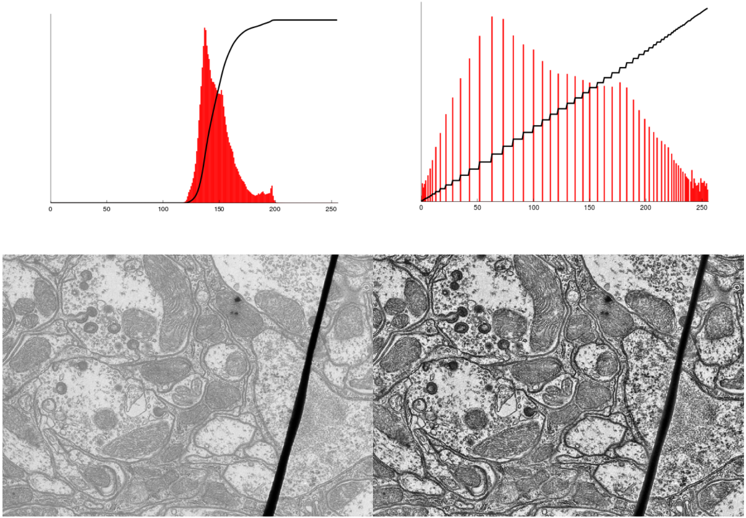
Deep Learning for Gender discrimination

Optimizer: Adam, Exponential LR scheduler





CHALE



References

References I

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. n.d. "Deep Residual Learning for Image Recognition."

Smith, Leslie N. 2015. "Cyclical Learning Rates for Training Neural Networks." *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*. Institute of Electrical; Electronics Engineers Inc., 464–72.

———. 2018. "A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay," March. http://arxiv.org/abs/1803.09820.

Smith, Leslie N., and Nicholay Topin. 2017. "Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates," August. SPIE-Intl Soc Optical Eng, 36. http://arxiv.org/abs/1708.07120.