Journal Report 13
1/6/20-1/10/20
Kevin Fu
Computer Systems Research Lab
Period 1, White

Daily Log

Monday January 6

Created empty/not-empty dataset from original, found accuracy suffered greatly. Tried increasing degree of data augmentation, didn't succeed. Switched to a piece-only dataset (removing empty), also low accuracy.

Tuesday January 7

Returned to previous data augmentations and tried training ResNet on piece-only dataset again. Achieved 84% training and 95% validation accuracy. Modified show_model.py to scrape highest accuracy from filenames with regex. Tried bumping up the validation split pct to get less noisy validation accuracy, which led to worse training.

Thursday January 9

Tried increasing and reducing learning rate, decided to keep at 1e-5. Played with epochs, found 200 epochs on a 152 layer ResNetV2 had slightly better training accuracy than before. Tried running ResNet on a white-piece only dataset, only able to achieve 87.5% validation accuracy.

Timeline

Date	Goal	Met
Dec 16	Write prediction-only script, combine with board-segmentation script to fulfill winter goal	Done
Winter Goal (Dec 19)	Have a script that converts a chessboard image to a digital array using a piece-recognition CNN (75% accuracy)	Done
Jan 6	Reorganize and comment code, figure out how to load Keras models quicker, consider training empty/not-empty network	Empty/not-empty tried
Jan 13	Continue experimenting for better piece-recognition model	Not started
Jan 20	Reintegrate new model into full code, organize and comment code, figure out quicker model loading for live demo	Not started

Reflection

I realized last (work)week that my CNN's accuracy was inflated because it overzealously marked empty squares. Because my dataset had more pictures of empty squares than squares with pieces, this behavior gave it better loss and accuracy, so it learned to guess empty too often. I tried turning it into a binary classification problem (empty square or not?) but the CNN isn't well suited to this task—an attention-based model, or even analyzing color changes between frames, would be much better. I also tried training a piece-only CNN, which again has a stellar validation accuracy (95%) but this is on a small sample size, which gives me pause.

I could try changing something more fundamental: reducing the height of each image from 2 (pose-estimated) squares to 1.75 or 1.5 squares might help, though I'm afraid this will cut off useful features of the image. Gathering more data would make it easier to train and test piece-only CNN, because I essentially halved my already-small dataset to train the current model.

Kevin Chung suggested sticking with my original model, but adding class weights to compensate for the imbalanced dataset. I'll also gather more images to expand my dataset. Hopefully, in tandem, these two changes will create a more robust piece-recognition model for the end-of-January demo.