11/8: results analysis

daily plan

- [x] look at experiment results and find insights into what params work best
- [x] run some more experiments
- [ ] build a "custom" cnn + fcnn combo and run it

results analysis

ok so takeaways:

- leakyrelu helps a lot
- for FC, batch norm by itself doesn't help. dropout helps a lot. both together is amazing
- opposite for CNN: batch norm is a substantial upgrade, dropout makes it worse (using dropout2d)
- using kernel size 5 for CNN looks best (probably because it considers the entire picture instead of just a random sub-picture of 3 players). also possible: kernel size 1 and then kernel size 5.
- lower learning rate seems better
- higher batch size seems better
- early stopping seems to have come up a LOT. hopefully lower LR will help with this, but it might be worthwhile disabling it.
- more complex FCs seemed worse

more experiments

deeper nets, higher batch sizes, lower learning rates. pretty much only using dropout+BN for FC, only dropout for ResNet, only BN for CNN. using leakyrelu instead of relu. CNN kernels size 5