Updating, re-training and refining your DeepLabCut model

Overfitting or generalizing, that is the question. What do you want to use your DLC model for? If you want to label many videos from a similar setup, an overfitted model will do very well (until you change the camera perspective, lighting conditions or maybe even the subject). On the other hand, if you wanted to train a model that detects human facial landmarks in different subjects, from different genders, ethnicities, facial hair styles and glasses and specially from different angles, then you need a model that generalizes well to all these conditions. And for that, you will need training examples, quite a few actually.

In this example we will start with a DLC model overfitted to detect facial landmarks, i.e., those that look exactly like mine. We will then try to analyze new videos from different subjects of different genders and ethnicities. We will extract outliers from these misdetections, refine them manually and re-train the model. Last, just for fun, I will use anipose to triangulate multiple camera perspectives into a final 3D facial expression.

1. Original model trained on 120 labeled frames of one single subject from 6 different camera angles

Figure 1: Labels of facial landmarks

1. DLC\_resnet50 with 1.030.000 iterations

Figure 2: Video of tracked face

1. Analyze 84 new videos from 5 different subjects
2. Extract outlier frames (numframes2pick: 5 because I was feeling lazy… and because 5x84 still is 420 frames to re-label manually…)
3. refine labels manually
4. merge dataset
5. train model again

The result is a DLC model trained on 540 frames from 6 different camera angles and 6 different subjects. Three females and three males, representing nationalities from Spain, Germany, Turkey, Iran, and China.

Figure 3: Collage of tracked facial expressions from different subjects

Extra. Recording with multiple camera angles has several benefits. First, different angles provide multiple training examples from the same recording, thus increasing your training set. Second, multiple angles avoid occlusion, e.g., when subjects roll their eyes behind your back, at least one camera will catch that. And last, multiple camera angles allow