

Etter å ha justert parametere, ville vi ha et mål på hvor god modellen var. For å få til dette, implementerte vi en metode som viste både treffsikkerhet for datasettet og en confusion matrix. Treffsikkerheten regner ut antall frames der modellen gjetter riktig kontra feil. Metoden trener først opp modellen på datasettet, og tester deretter på en ny, separert del av datasettet. Dette gjøres for å unngå overtilpasning der modellen “husker” bildene i datasettet. Resultatet måles i prosent, og ble i vårt tilfelle rundt 95%. Confusion matrixen viser hvordan treffsikkerheten er fordelt over de forskjellige skiltene i datasettet. Den viser label for hver rad og prediksjon for hver kolonne. Skiltene er sortert slik: [ukjent, 20, 30, 40, 50, 60, 70, 80, 100, 120]. Her ser vi stort sett verdier langs diagonalen, som viser god nøyaktighet.

accuracy: 93.56%

confusion matrix:

```
[[296  0  0  0  0  0  0  0  0  0]
 [  0 19  0  0  0  0  0  0  0  0]
 [  2  0 202  0  8  0  0  2  0  0]
 [  1  0  0 116  0  0  0  0  0  0]
 [  0  0 19  0 213  2  0  9  1  0]
 [  1  0  0  0  2 125  0 10  0  0]
 [  0  0  0  0  0  0 181  0  0  1]
 [  1  0  3  0  4 10  0 154  4  1]
 [  0  0  1  0  2  1  0  4 124  4]
 [  0  0  0  0  4  0  0  2  7 110]]
```

I tillegg gjorde vi det samme for nøyaktigheten for et tenkt produksjonsmiljø. I dette tilfellet tok vi en video modellen ikke hadde sett før, og lagra alle skiltene i videoen med forventet label. Etter litt manuelt arbeid med å korrigere riktig label, kunne vi da kjøre modellen på nytt med dette test-datasettet, og få nøyaktigheten på samme måte som med funksjonen vi brukte på testdatasettet.

confusion matrix:

```
[[ 1  0  0  0  0  0  0  0  0  0]
 [ 0  1  0  0  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0]]
```

```
[ 0 1 0 65 0 0 0 0 0 0]
[ 0 1 32 0 106 0 0 1 0 0]
[ 1 0 27 3 40 0 0 0 0 0]
[ 0 10 10 16 3 0 36 2 0 0]
[ 0 0 0 0 0 0 0 1 0 0]
[ 0 0 0 0 0 0 0 0 1 0]
[ 0 0 0 0 0 0 0 0 0 1]]
```

Når dette var gjort, fikk vi testet 40-, 50-, 60-, og 70-skilt, og vi la merke til at 60- og 70-skiltene hadde mye dårligere treffsikkerhet i dette tilfellet. Dette mistenkte vi å ha med datasettet å gjøre fordi bildene for de skiltene ikke var like bra som resten.

Her ser vi altså nøyaktigheten for hvert skilt. 50-skilt blir av og til klassifisert som 30. 60-skilt tolkes feil hver gang, og 70-skilt tolkes kun riktig omtrent halvparten av gangene. Vi dro derfor ut igjen for å samle inn mer video av skiltene som gjorde det dårligst for å utvide datasettet med bedre bilder.

Når vi fikk samlet inn nok skilter til å utvide datasettet, ble resultatet mye bedre: Vi ser nå at 50 og 60-skilt blir forvekslet med hverandre omtrent 1/10 ganger. Grunnen til at 20, 30, 40, 80, og 120 mangler her er fordi vi ikke rakk å samle inn nok testdata for disse. Disse har allikevel blitt brukt under trening, og mangler kun testing. I og med at dataen allerede har blitt brukt for trening, vil testing med den samme dataen gi en overdreven nøyaktighet grunnet overtilpasning, og var derfor ikke aktuelt.

accuracy: 86.31 %

confusion matrix:

```
[[ 1 0 0 0 0 0 0 0 0 0]
 [ 0 0 0 1 0 0 0 0 0 0]
 [ 0 0 1 0 0 0 0 0 0 0]
 [ 0 1 0 65 0 0 0 0 0 0]
 [ 1 1 7 0 109 11 0 5 0 0]
 [ 0 0 1 0 7 63 0 0 0 0]
 [ 1 1 4 0 4 0 69 2 0 0]]
```

```
[ 0 0 0 0 0 0 0 0 1 0 0]
[ 0 0 0 1 0 0 0 0 0 0 0]
[ 0 0 0 0 1 0 0 0 0 0 0]]
```