**Agenda meeting Maaike**

**Datum: 23 maart 2016**

**1. Wat hebben we deze week gedaan?**

**Taken afgelopen week:**

Daniëlle: jobscript batches uitgezocht

Laurens: begin gemaakt aan k-means en mean shift

Roos: caffemodel feature extraction aan de praat krijgen

Diede: caffemodel feature extraction op de server draaien

**Resultaten:**

- mean shift maakt maar 1 cluster aan, op testset van 137 featurevectoren. Waarom?

- feature extraction op de server draaien werkt, maar duurt erg lang. 84 uur voor alle train images. Kan het efficiënter?

**Woensdag:**

- Daniëlle: SVM implementeren (nog op basis van testset van 137 featurevectoren)

- Roos: SVM oriënteren + data bekijken + ziek naar huis

- Laurens: k-means & mean shift. Waarom maar 1 cluster? hoe gaat caffe om met te kleine/te grote plaatjes? welke images worden bij de medium images verwijderd en wat zijn hun kenmerken?

- Diede: kijken of feature extraction op de server efficiënter kan. Op scratch-space images gedraaid. Batch size aangepast. Subdatasets maken.

**Resultaten:**

- mean shift creert nog steeds maar 1 cluster.

- feature extraction zou nu in 10 losse jobs gerund kunnen worden.

**Vragen:**

- We zijn nu van plan om de medium plaatjes te gebruiken om de feature vectoren mee te genereren. Deze zijn 250x250. Is dit slim? Caffemodel verwacht input van 224x224 (denken we), en we zijn bang dat kleinere plaatjes gezeropad worden, wat niet goed zou zijn voor de features.

*Weggooi statistieken laten zien. Hoeveel foto’s er dan weggegooid worden: ongeveer 1 procent. Redelijk verspreid over de restaurants.*

*Caffenet: Semi random in het midden het plaatje nemen. 250x250 is dus prima.*

*Mean shift: maar 1 cluster gevonden. Is mean shift beter of k-means? Mean shift: handig dat hij clusters vindt. Maar: dan moet hij er wel meerdere vinden. Bij k-means kun je zelf instellen hoeveel clusters. Mean shift sloom op grote datasets.*

*Rond de 20 clusters? Maaike: of misschien wel 100. Je hebt liever dat meerdere clusters voor 1 label zorgen. Dus kijken of mean shift snel genoeg is, en hoeveel clusters hij dan maakt.*

*Hoe efficiënter maken van feature extraction.*

*Erdi vragen: hoe hebben jullie dat gedaan met de features? Aantal uren? Qua batch job scripts, hoeveel nodes, hoeveel cores.*

*Mensen mailen surfsara hoe gaan de uren zo snel op, uren per core?*

*Vandaag: features extraction aanzetten! Zodat we in het weekend aan het werk kunnen.*

*Vraag Roos: Bedrijven met heel weinig foto’s: wat wil je daarmee? Puur op basis van weinig fotos laten trainen, werkt dat?*

*Een andere optie is om die bedrijven niet mee te laten trainen, en dan de bedrijven op kans te classificeren.*

*We kunnen het beiden proberen, kijken welke beter werkt. Gewoon 2 submissions.*

**Wat zijn onze taken deze week?**

Aanname: zaterdag zijn de train feature vectoren er.

Daniëlle:

- SVM toepassen op featurevectoren.

- Andere dingen. Weka, sklearn andere methodes.

Roos:

- beter worden.

- random forest toepassen op train feature vectoren.

Laurens:

- cluster toepassen op de feature vectoren.

- stap 2: prototypes gebruiken om te binnen.

- Erdi mailen/facebook.

Diede:

- runnen van de feature vectoren. Zaterdag de train featurevectoren af. Maandag de test featurevectoren af.

- surfsara mailen

**Wat is onze verdere planning?**

Woensdag:

- onze eerste submission, op basis van SVM direct op featurevectoren.

- werken aan het gehele model: clusters. 🡪 meteen tweede submission?

Hopelijk: ensemble maken van het Directe model, woordenboek. Op basis van de kansen die de modellen aan de labels geven.