

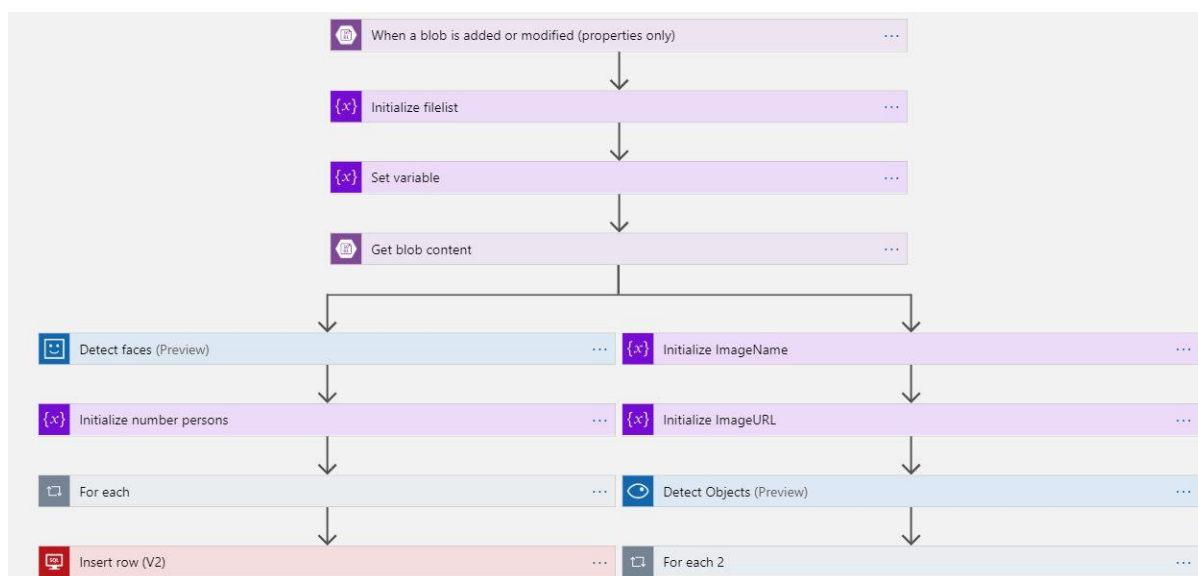
## People counting på Århus Universitet

Århus Universitet (AU) har et ønske om at udvide funktionaliteten af overvågningskameraerne i auditorierne, således at de også kan bruges til at tælle antallet af tilstedeværende personer på forskellige tidspunkter af dagen.

AU bruger i forvejen en del services i Azure, og vil derfor gerne undersøge om Cognitive Services i Azure kan bruges til formålet. Cognitive services dækker over en lang række færdigpakke funktioner indenfor kunstig intelligens, og herunder også computer vision, som er den relevante funktion i dette tilfælde. Vi har undersøgt to forskellige API'er indenfor computer vision:

- 1) Faces API, som er beregnet til ansigtsgenkendelse, registrering af ansigtsudtryk, og som også kan bruges til en simpel optælling af antallet af ansigter på billedet.
- 2) Object detection API, som er beregnet til at genkende en lang række forskellige objekter på billedet, herunder personer.

Indledningsvis har vi manuelt opsamlet 20 billeder fra 3 forskellige kameraer placeret i 3 forskellige auditorier. Til analysen af billederne med de to API'er har vi bygget en Logic App, som vist på figuren herunder.



Figur 1 Logic App til kald af Cognitive Services

App'en bliver triggeret hver gang der tilføjes et nyt billede til en blob. Filnavn og filindhold udtrækkes, og herefter deler app'en sig i to. I venstre side kaldes Face API'et, og ud fra resultatet tælles antallet af ansigter, og dette antal skrives sammen med filnavnet til en tabel i en database. I højre side kaldes i stedet Object Detection API'et. Resultatet indeholder alle de objekter som API'et kan finde på billedet, og de skrives alle sammen til en tabel i en database sammen med filnavnet. Herefter filtreres på objekter af typen "person".

Resultatet fra de 20 billeder ses i tabellen herunder. Kolonnerne Faces\_Antal og Objects\_Antal indeholder resultatet fra de to API'er. Kolonnen Ground\_truth indeholder det faktiske antal personer på billedet, bestemt ved manuel optælling. Det skal bemærkes, at det bliver sværere og sværere at bestemme det faktiske antal personer, jo flere personer der er på billedet. Kamera ACCC8E61BCD8 sidder i et stort auditorie, og når der er mere end 200 personer tilstede er usikkerheden af størrelsen +/- 10 personer.

	Filename	Faces_Antal	Objects_Antal	Camera	Ground_truth
1	ACCC8E41737E-2020-03-02_06-02-00Z-0.jpg	0	0	ACCC8E41737E	0
2	ACCC8E41737E-2020-03-04_12-41-00Z-13.jpg	2	3	ACCC8E41737E	13
3	ACCC8E41737E-2020-03-02_12-48-00Z-20.jpg	4	3	ACCC8E41737E	20
4	ACCC8E41737E-2020-03-04_11-34-00Z-26.jpg	10	8	ACCC8E41737E	26
5	ACCC8E41737E-2020-03-04_12-51-00Z-26.jpg	10	7	ACCC8E41737E	26
6	ACCC8E41737E-2020-03-02_09-30-00Z-40.jpg	32	4	ACCC8E41737E	40
7	ACCC8E41737E-2020-03-04_11-12-00Z-23.jpg	12	2	ACCC8E41737E	23
8	ACCC8E41737E-2020-03-04_10-51-00Z-2.jpg	0	2	ACCC8E41737E	2
9	ACCC8E44A4AE-2020-03-02_06-02-00Z-0.jpg	0	0	ACCC8E44A4AE	0
10	ACCC8E44A4AE-2020-03-02_12-48-00Z-35.jpg	0	1	ACCC8E44A4AE	35
11	ACCC8E44A4AE-2020-03-02_09-30-00Z-0.jpg	0	0	ACCC8E44A4AE	0
12	ACCC8E61BCD8-2020-03-02_09-30-00Z-278.jpeg	0	0	ACCC8E61BCD8	278
13	ACCC8E61BCD8-2020-03-02_06-02-00Z-1.jpeg	0	0	ACCC8E61BCD8	1
14	ACCC8E61BCD8-2020-03-04_11-10-00Z-118.jpg	0	0	ACCC8E61BCD8	118
15	ACCC8E61BCD8-2020-03-04_08-58-00Z-200_plus.jpg	0	0	ACCC8E61BCD8	200
16	ACCC8E61BCD8-2020-03-04_10-51-00Z-200_plus.jpg	0	0	ACCC8E61BCD8	200
17	ACCC8E61BCD8-2020-03-04_12-41-00Z-150_plus.jpg	0	0	ACCC8E61BCD8	150
18	ACCC8E61BCD8-2020-03-04_09-10-00Z-176.jpg	0	0	ACCC8E61BCD8	176
19	ACCC8E61BCD8-2020-03-02_12-48-00Z-200_plus.jpeg	0	1	ACCC8E61BCD8	200
20	ACCC8E61BCD8-2020-03-04_11-34-00Z-150_plus.jpg	0	1	ACCC8E61BCD8	150

Figur 2 Resultat af Faces API og Object Detection API på 20 billeder

Når man kigger ned over tallene ses, at Faces API'et ser ud til at være mere velegnet til formålet end Object Detection API'et. Men samtidig ses, at begge API'er kommer til kort i det store auditorie med mange mennesker. Det foreliggende antal billeder er dog for lille til at kunne sige noget om hvor grænsen går for, hvornår man kan bruge Faces API'et.

I lyset af den manglende funktionsevne for begge API'er i det store auditorie har vi undersøgt andre muligheder. Det viser sig, at der findes et begreb, der hedder Crowd Counting, som går ud på at bestemme antallet af mennesker i store forsamlinger, som f.eks demonstrationer og sportsbegivenheder. Og der er også udviklet machine learning modeller til formålet. Det er beskrevet flere steder, som f.eks i [denne artikel](#). I artiklen omtales en model, der hedder CSRNet. Denne model findes i to udgaver, beregnet til hhv. tætte folkemængder (model A) og knapt så tætte folkemængder (model B). Herunder vises billeder, som er repræsentative for de to udgaver af modellen.



Figur 3 Repræsentativt billede for CSRNet model A (tætte folkemængder)



Figur 4 Repræsentativt billede for CSRNet model B (mindre tætte folkemængder)

De to modeller er implementeret som et Python script på en virtuel maskine i Azure. Dette script vil kunne implementeres som en Function App, og kunne triggres hver gang et billede bliver tilføjet til en blob.

I tabellen herunder er der tilføjet to kolonner med resultatet fra CSRNet model A og B. Det ses, at disse to modeller begge giver et resultat for det store auditorium (camera ACCC8E61BCD8), og umiddelbart ser model A ud til at være tættest på det rigtige resultat. Det ses også, at de to CSRNet modeller begge overvurderer antallet, når der kun er få personer tilstede.

	Filename	CSR_Antal_A	CSR_Antal_B	Faces_Antal	Objects_Antal	Camera	Ground_truth
1	ACCC8E41737E-2020-03-02_06-02-00Z-0.jpg	37	26	0	0	ACCC8E41737E	0
2	ACCC8E41737E-2020-03-04_12-41-00Z-13.jpg	72	50	2	3	ACCC8E41737E	13
3	ACCC8E41737E-2020-03-02_12-48-00Z-20.jpg	80	47	4	3	ACCC8E41737E	20
4	ACCC8E41737E-2020-03-04_11-34-00Z-26.jpg	71	27	10	8	ACCC8E41737E	26
5	ACCC8E41737E-2020-03-04_12-51-00Z-26.jpg	105	53	10	7	ACCC8E41737E	26
6	ACCC8E41737E-2020-03-02_09-30-00Z-40.jpg	123	110	32	4	ACCC8E41737E	40
7	ACCC8E41737E-2020-03-04_11-12-00Z-23.jpg	81	59	12	2	ACCC8E41737E	23
8	ACCC8E41737E-2020-03-04_10-51-00Z-2.jpg	38	21	0	2	ACCC8E41737E	2
9	ACCC8E44A4AE-2020-03-02_06-02-00Z-0.jpg	25	13	0	0	ACCC8E44A4AE	0
10	ACCC8E44A4AE-2020-03-02_12-48-00Z-35.jpg	61	52	0	1	ACCC8E44A4AE	35
11	ACCC8E44A4AE-2020-03-02_09-30-00Z-0.jpg	30	32	0	0	ACCC8E44A4AE	0
12	ACCC8E61BCD8-2020-03-02_09-30-00Z-278.jpeg	351	359	0	0	ACCC8E61BCD8	278
13	ACCC8E61BCD8-2020-03-02_06-02-00Z-1.jpeg	27	53	0	0	ACCC8E61BCD8	1
14	ACCC8E61BCD8-2020-03-04_11-10-00Z-118.jpg	182	235	0	0	ACCC8E61BCD8	118
15	ACCC8E61BCD8-2020-03-04_08-58-00Z-200_plus.jpg	363	371	0	0	ACCC8E61BCD8	200
16	ACCC8E61BCD8-2020-03-04_10-51-00Z-200_plus.jpg	272	304	0	0	ACCC8E61BCD8	200
17	ACCC8E61BCD8-2020-03-04_12-41-00Z-150_plus.jpg	286	308	0	0	ACCC8E61BCD8	150
18	ACCC8E61BCD8-2020-03-04_09-10-00Z-176.jpg	176	238	0	0	ACCC8E61BCD8	176
19	ACCC8E61BCD8-2020-03-02_12-48-00Z-200_plus.jpeg	356	353	0	1	ACCC8E61BCD8	200
20	ACCC8E61BCD8-2020-03-04_11-34-00Z-150_plus.jpg	287	299	0	1	ACCC8E61BCD8	150

Figur 5 Resultat af CSR model A, CSR model B, Faces API og Object Detection API på 20 billeder

Afslutningsvis skal det bemærkes at disse resultater bygger på et meget spinkelt datagrundlag på bare 20 billeder fra 3 auditorier. Tidligere i denne uge er det lykkedes at få automatiseret processen med at uploade billeder fra kameraerne, og når de studerende vender tilbage efter corona pausen forventer vi at opsamle et mere omfattende billedmateriale. Det har også vist sig, at nogle af kameraerne ikke er justeret optimalt i forhold til at give den bedst mulige dækning af auditorierne, og det forventer vi også at rette op på. Tilsammen skulle det gerne give mulighed for at udarbejde en egentlig statistik på hvor god metoden er.