Unsupervised deep learning based video

Luca Barras,
257916

Objectives and motivations

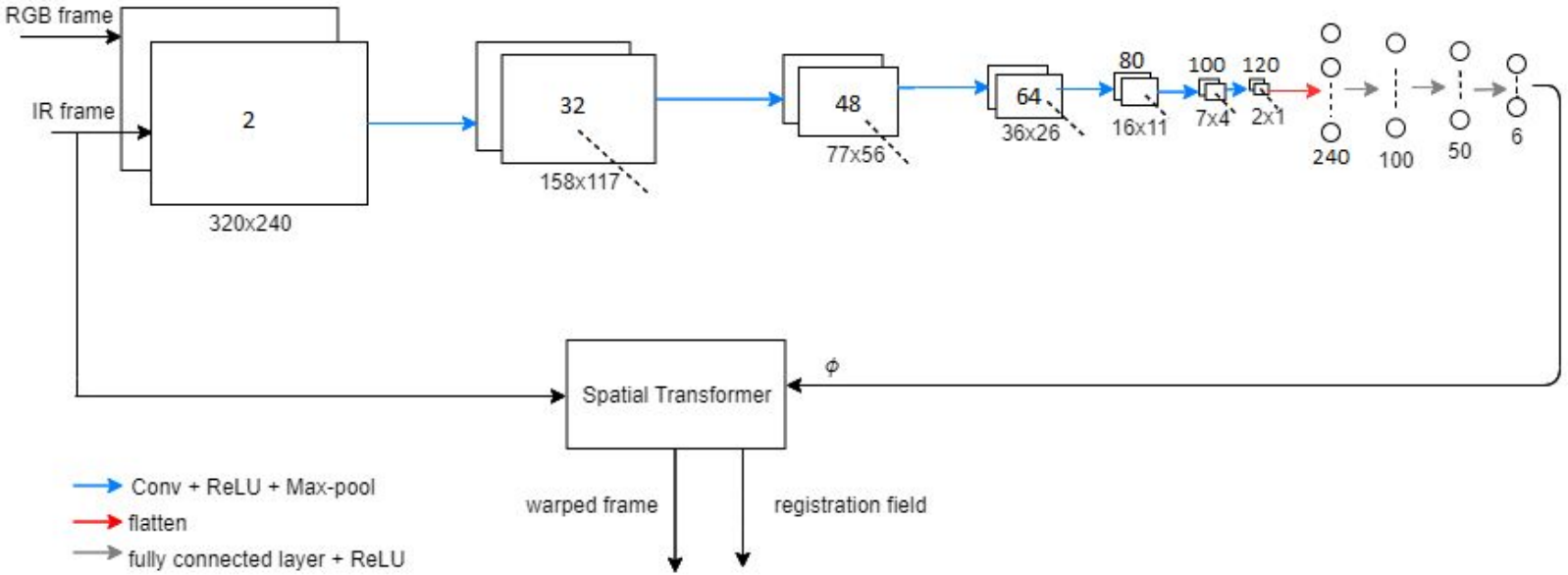
- ➔ Registration between an Infrared (IR) and classical (RGB) video
- ➔ IR camera is sensible to the heat
- ➔ Some methods to align IR and RGB images but not with lots of data (videos)
- ➔ Unsupervised deep learning method

Example of a pair of RGB and IR videos

RGB video

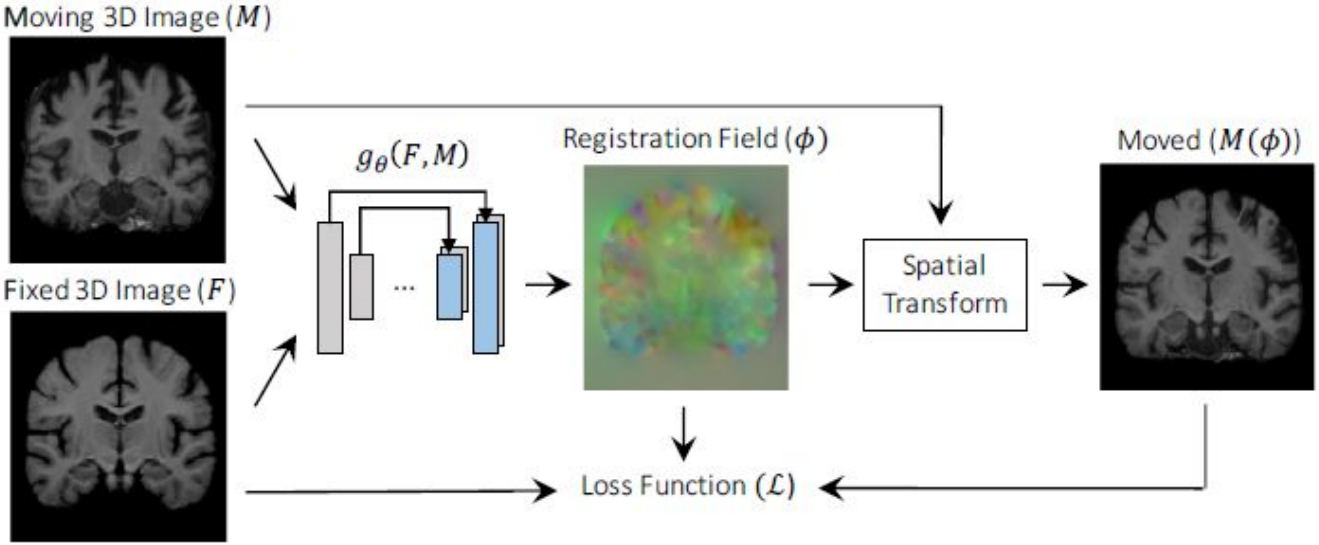
IR video

Affine network



$$\mathcal{L}(F, M, \phi) = \mathcal{L}_{sim}(F, M(\phi))$$

Voxelmorph network

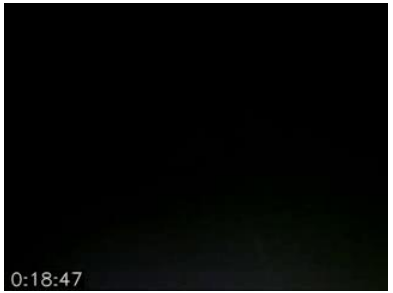


$$\mathcal{L}(F, M, \phi) = \mathcal{L}_{sim}(F, M(\phi)) + \lambda \mathcal{L}_{smooth}(\phi)$$

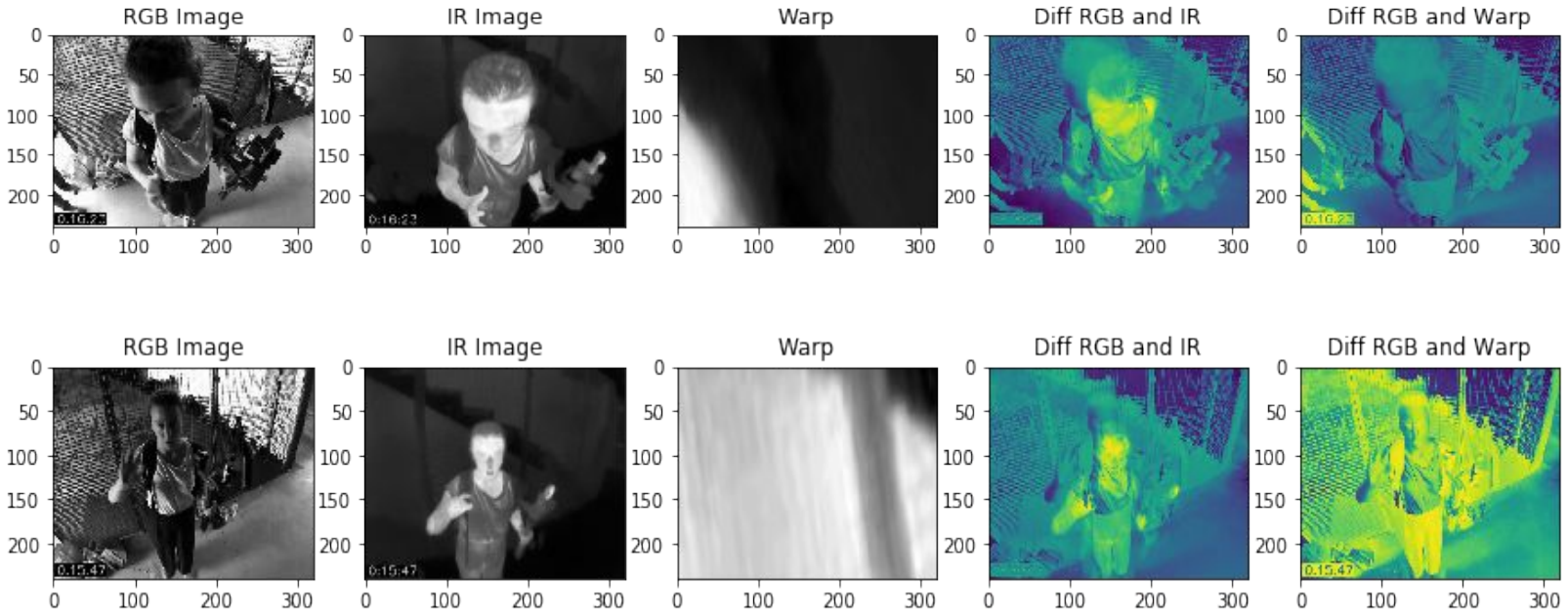
https://arxiv.org/abs/1802.02604

Dataset (1st part)

- ➔ A serie of RGB and IR videos pairs
- ➔ Preprocessing on the dataset:
  - ◆ Convert videos in frames
  - ◆ Resize IR frames
  - ◆ Remove too dark frames
  - ◆ Use the gray-scale version of the frames

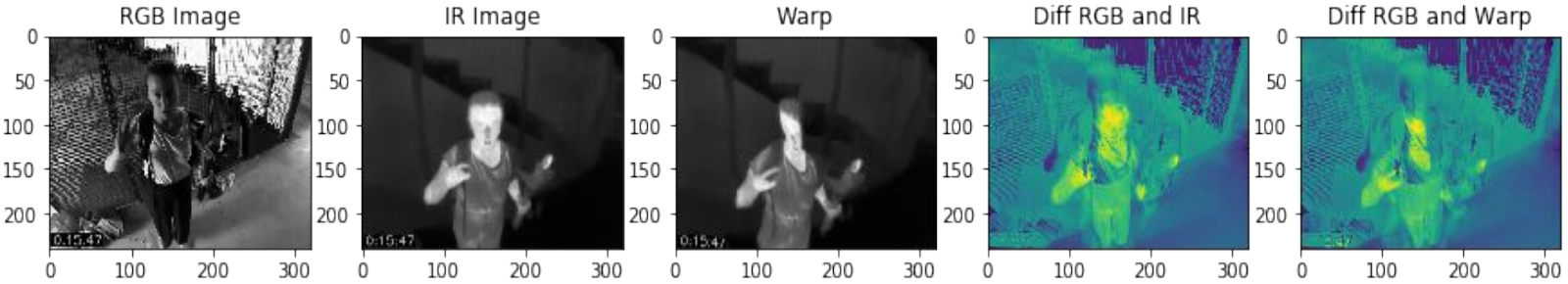


Results (1st part) Affine network

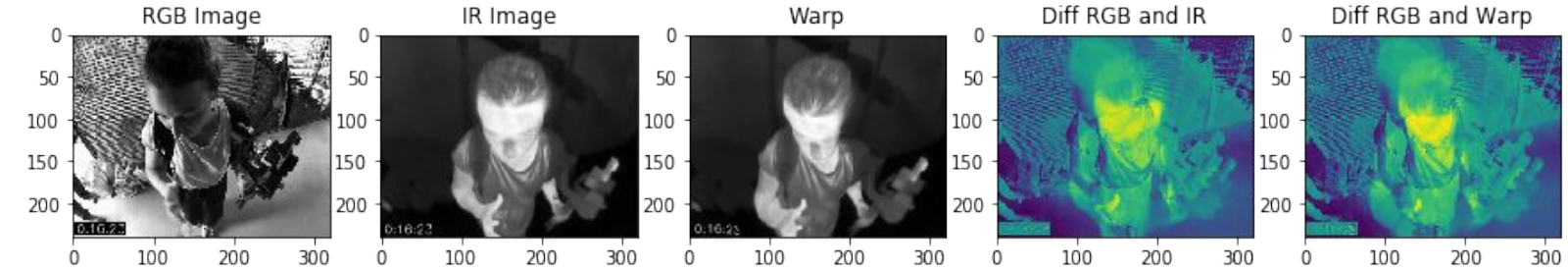


Results (1st part) Voxelmorph network

λ = 0.2



λ = 0.6



Results analysis (1st part)

- ➔ In affine network, results too scale up and not align
- ➔ In Voxelmorph network, results too warped and not align
- ➔ No correlation between the pixel values of the RGB and IR frames

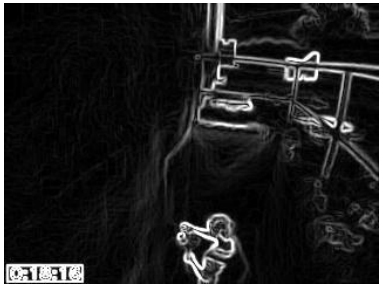
Dataset (2nd part)

- ➔ We reuse exactly the same dataset as in the 1st part (gray-scale version)
- ➔ We apply Sobel, Canny, adaptive threshold and laplacian filter

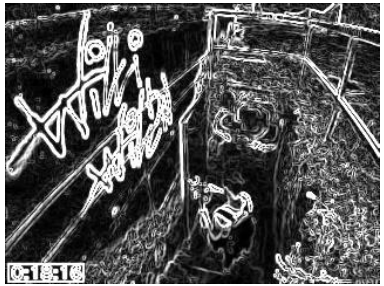
Filter results

Sobel

IR



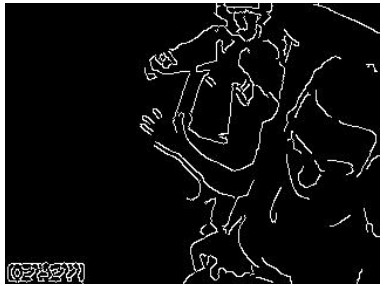
RGB



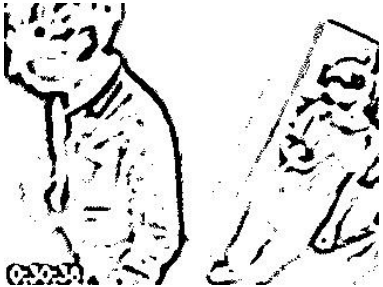
Canny

IR

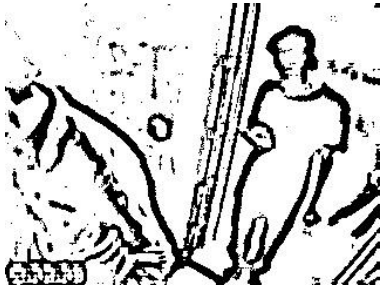
RGB



Adaptive threshold



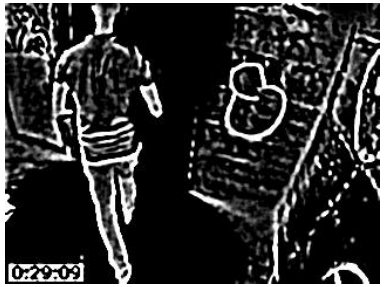
IR



RGB

Laplacian

IR



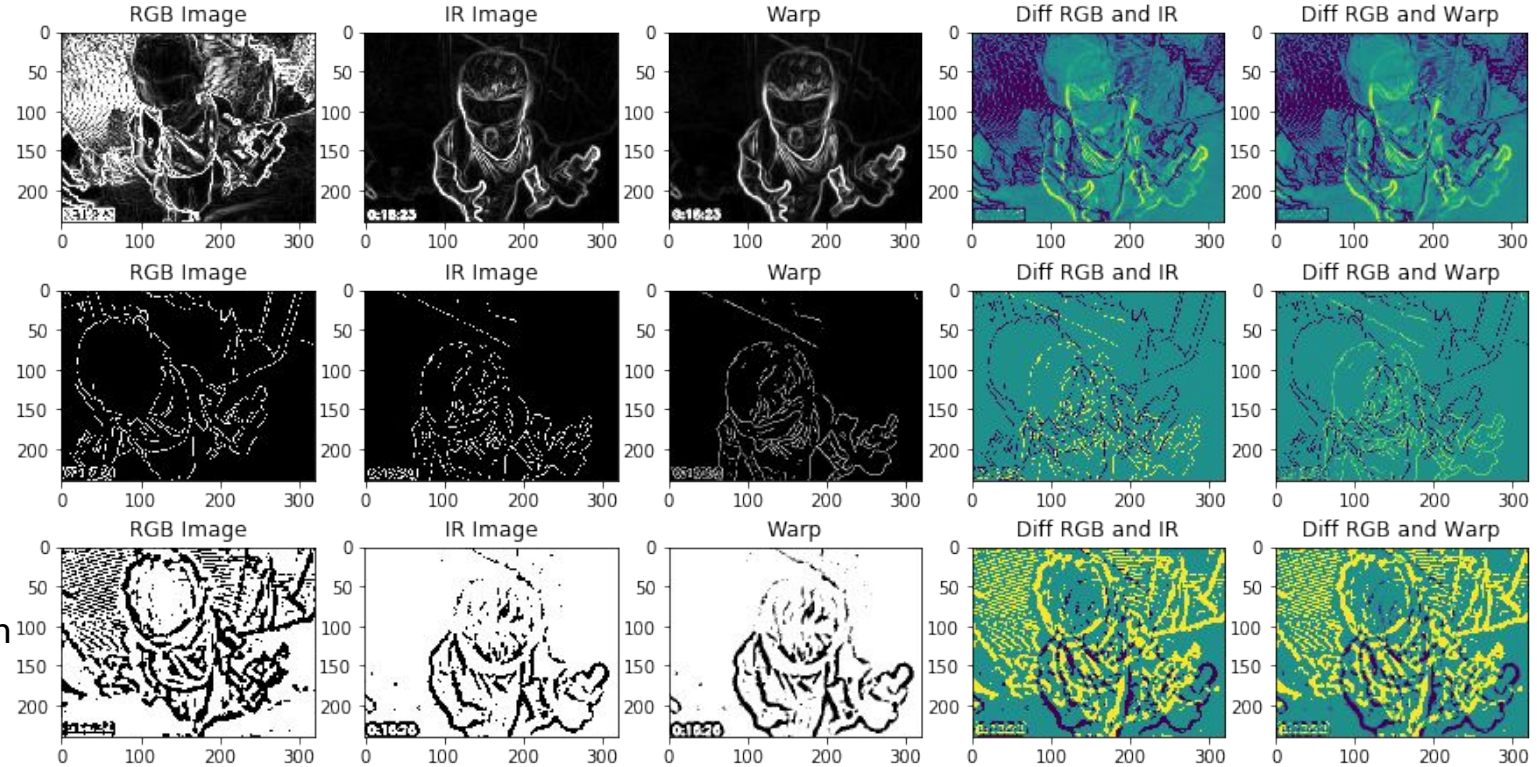
RGB

Results (2nd part) Affine network

Sobel

Canny

Ad. thresh

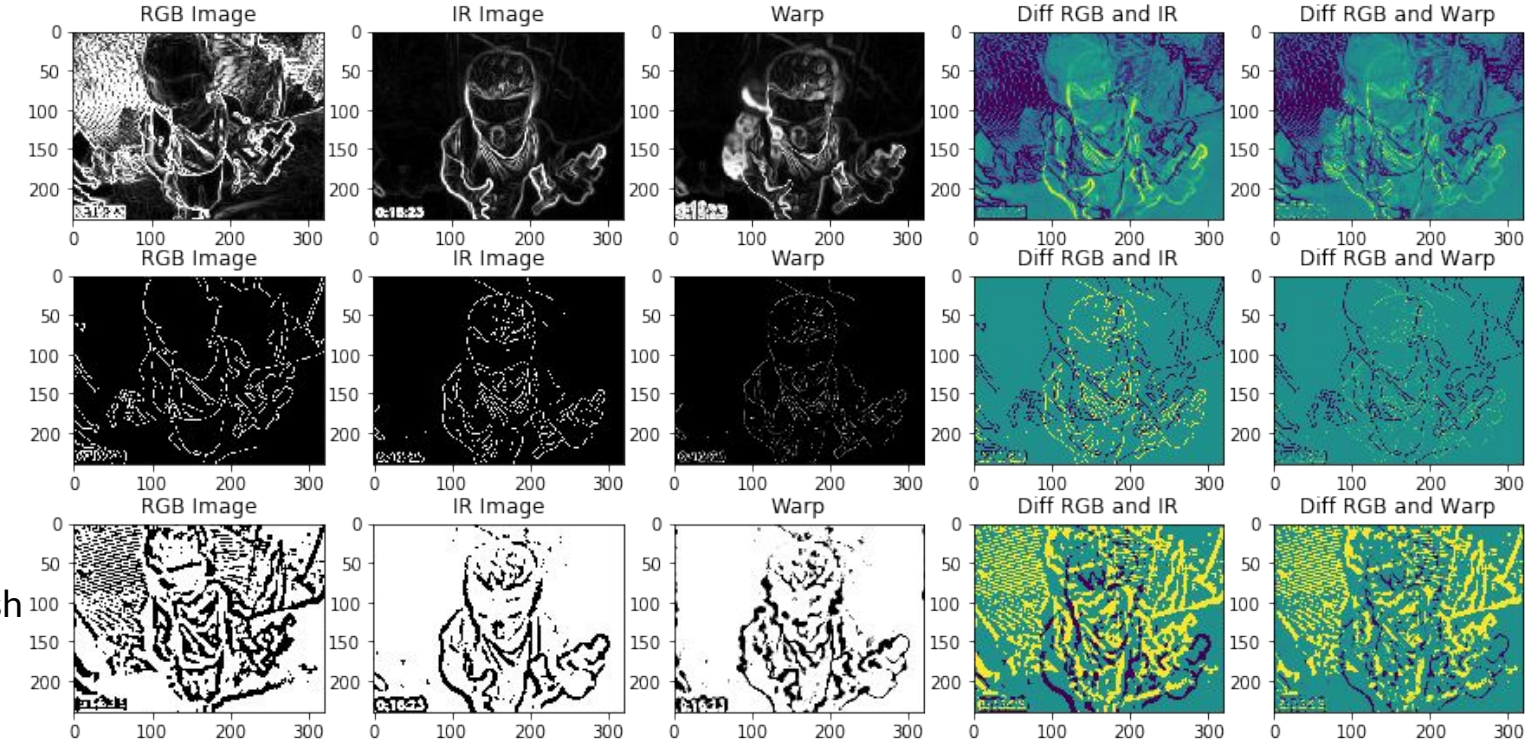


Results (2nd part) Voxelmorph network

Sobel
λ = 0.2

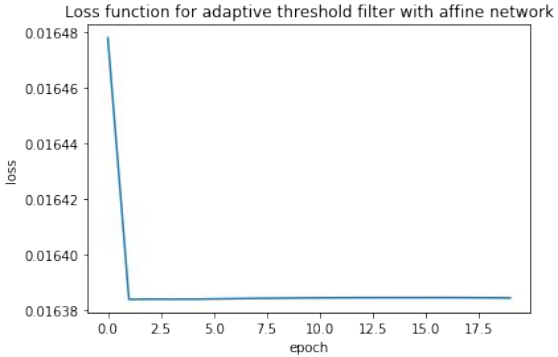
Canny
λ = 0.6

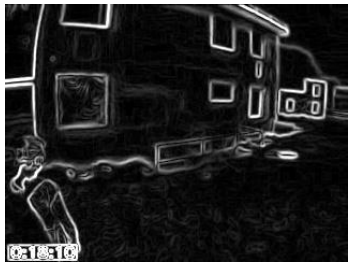
Ad. thresh
λ = 0.2



Results analysis (2nd part)

- ➔ In Affine network, the result does not move at all
- ➔ In Voxelmorph netwok, the results move only locally (deform shape)
- ➔ Loss function is too flat and much more contrast on RGB frames than IR frames





IR frame

RGB frame

Dataset (3rd part)

1. For each video, we keep only frames where are people
2. Use Deeplab to segment the RGB frames
3. Threshold with the blue channel to segment IR frames (Adapt thresh. depending of the sequence of frames)
4. Keep only frame we can well segment (IR and RGB)
5. Finally we have 10431 4-tuple of frames (8000 training set, 1000 validation set, 1400 test set)

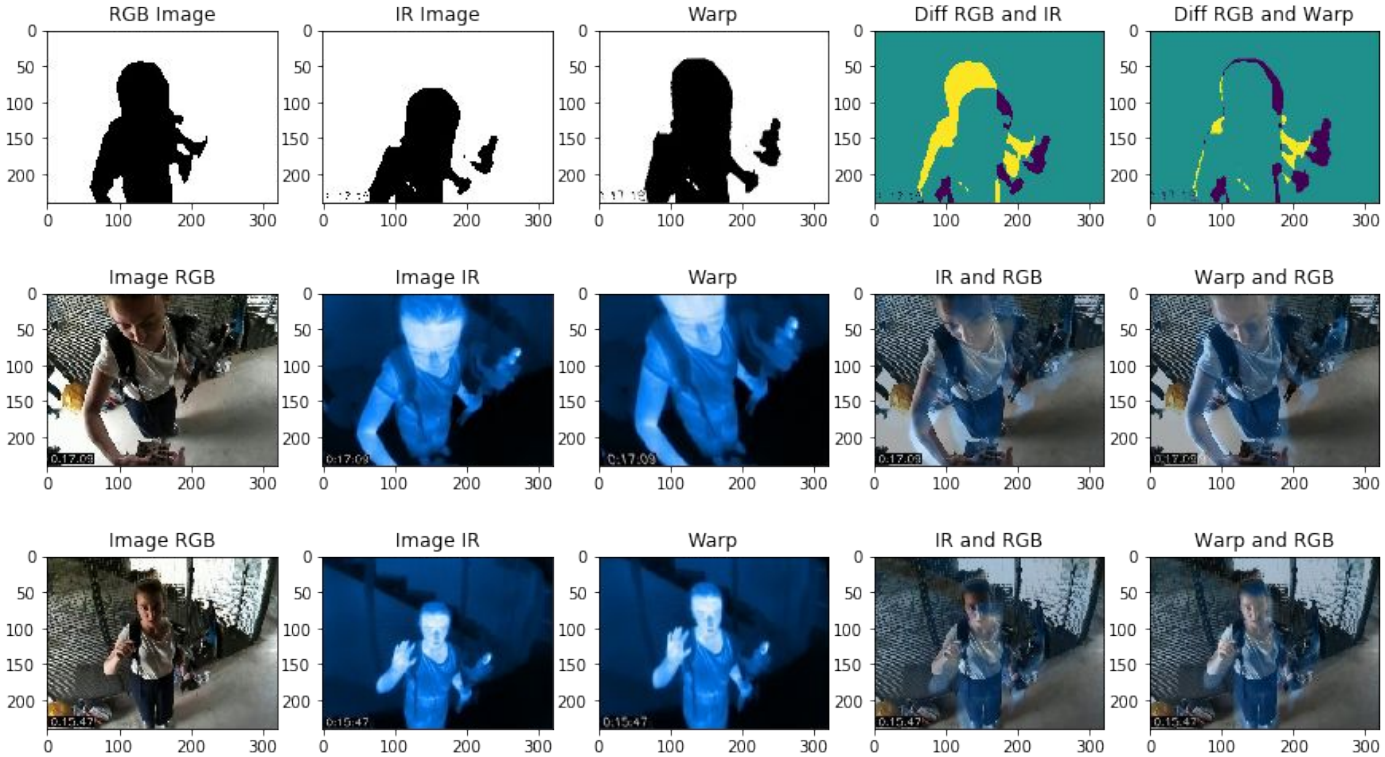
RGB

IR

mask RGB

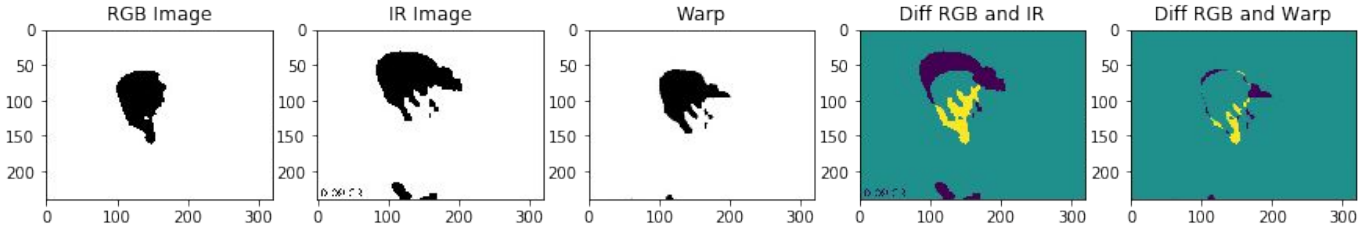
mask IR

Results (3rd part) Affine network

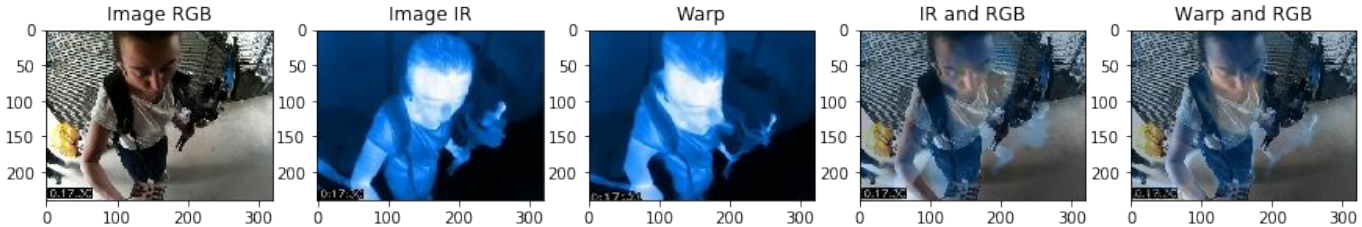


Results (3rd part) Voxelmorph network

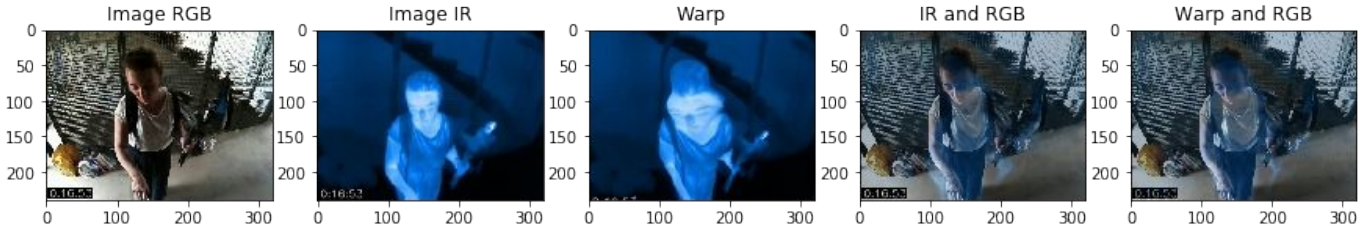
mask
λ = 1.0



λ = 0.6



λ = 1.0

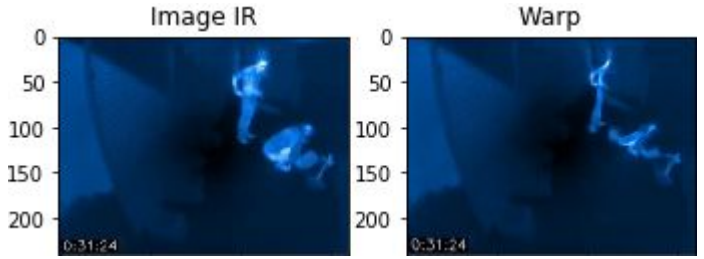


Results analysis (3rd part)

- ➔ We have pretty good results but not again perfect
- ➔ The segmentation could be better (RGB and IR)
- ➔ Problem when two frames are too far each others (intersection between RGB and IR mask empty)

RGB mask

IR mask



Results Affine network

Before registration

After registration

Conclusion

- ➔ Registration between IR and RGB frames are not as obvious
- ➔ We must find a correlation
- ➔ Contours is a correlation but loss function too flat and RGB frames much more contrast than IR frames
- ➔ Segmentation is better but not again perfect (segmentation precision could be better)
- ➔ For better results, we must improve the segmentation step

https://github.com/lbarras27/deeplearning_registration_IR_RGB_video

Thanks for your attention