

LA DETECTION DE FLUX OPTIQUE (AVEC RAFT)



LA DÉTECTION DE FLUX OPTIQUE

La detection de flux optique consiste à estimer la différence de mouvement par pixel entre deux images.

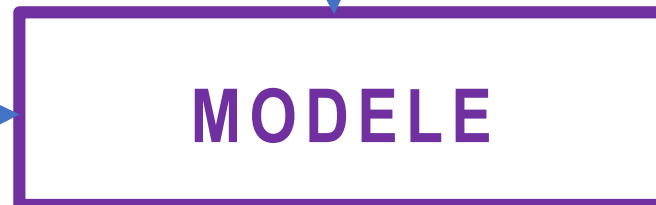
► **Pas tout à fait équivalent à de la détection de mouvement**

- un flux optique peut venir du mouvement de la caméra.
- Equivalents si la caméra est immobile

► Tache de **segmentation** d'image

- Entrée image(s), sortie image

Frame ou informations de
la frame précédente

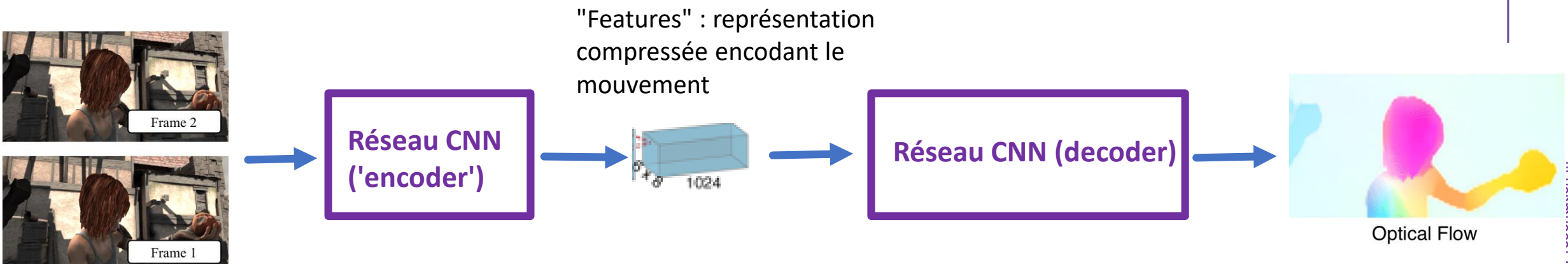


Optical Flow

FLOWNET

L'ancêtre de RAFT (2020)

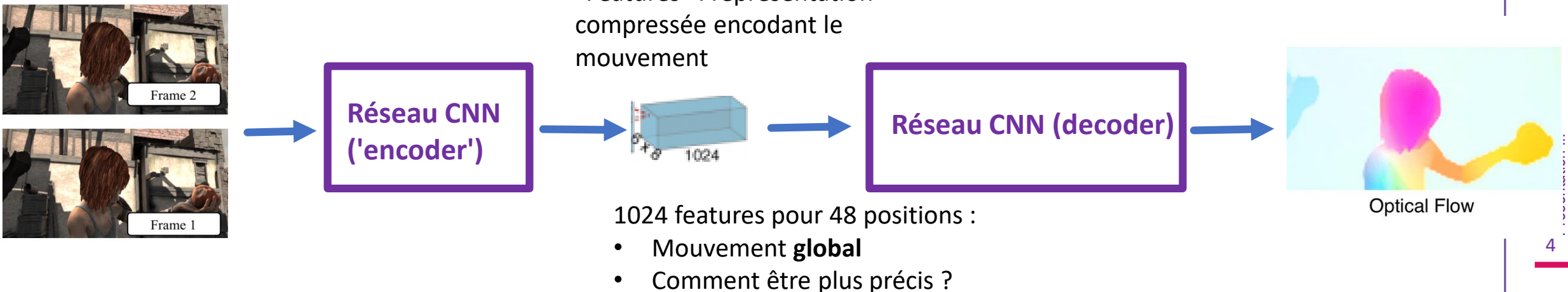
- ▶ 2015
- ▶ En entrée : deux images concaténées -> une "image" de 6 channels R1,G1,B1,R2,G2,B2, .
- ▶ Composé de réseau de convolution simples proche d'un encoder/decoder
- ▶ Apprend à sortir le flux optique entre les deux images



FLOWNET

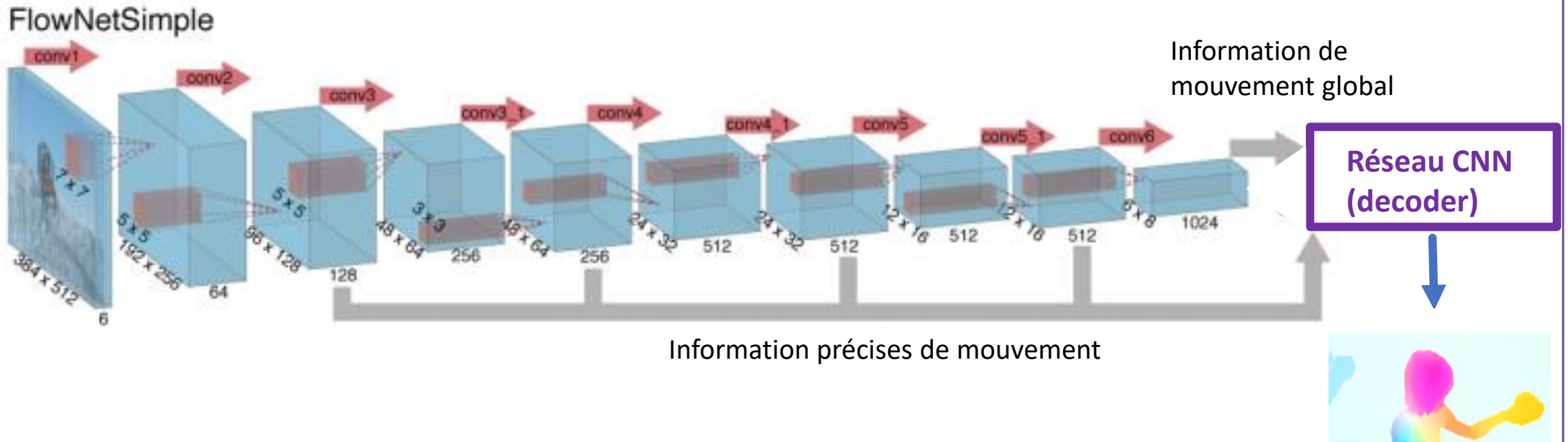
L'ancêtre de RAFT (2020)

- ▶ 2015
- ▶ En entrée : deux images concaténées -> une "image" de 6 channels R1,G1,B1,R2,G2,B2, .
- ▶ Composé de réseau de convolution simples proche d'un encoder/decoder
- ▶ Apprend à sortir le flux optique entre les deux images



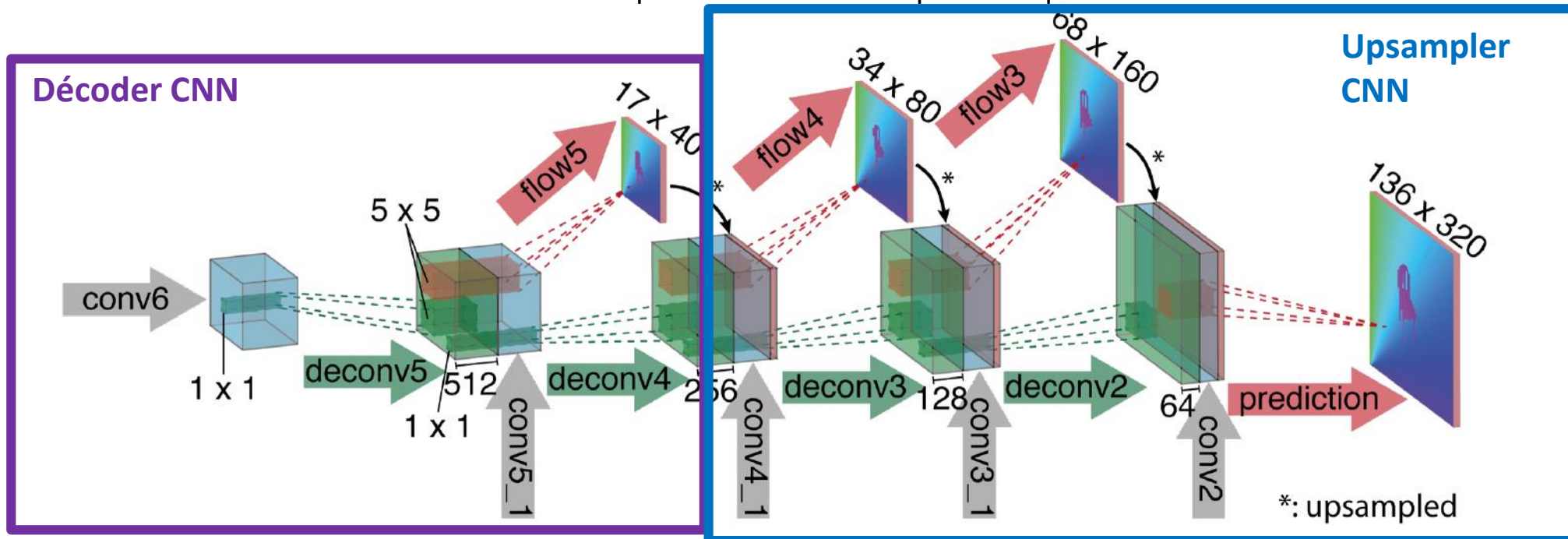
FLOWNET : ENCODER

- Solution dans Flownet: utiliser les propriétés du CNN pour envoyer à la fois des informations de mouvement globale et précises au decoder
- Le decoder apprend à générer une image de mouvement précise à **partir du mouvement global (sortie de l'encoder) et d'informations précises** (couches inférieures).
- Pour faire simple, il "augmente la résolution" de la sortie de l'encoder, passant d'une image 6*8 à une image 136 x 320.



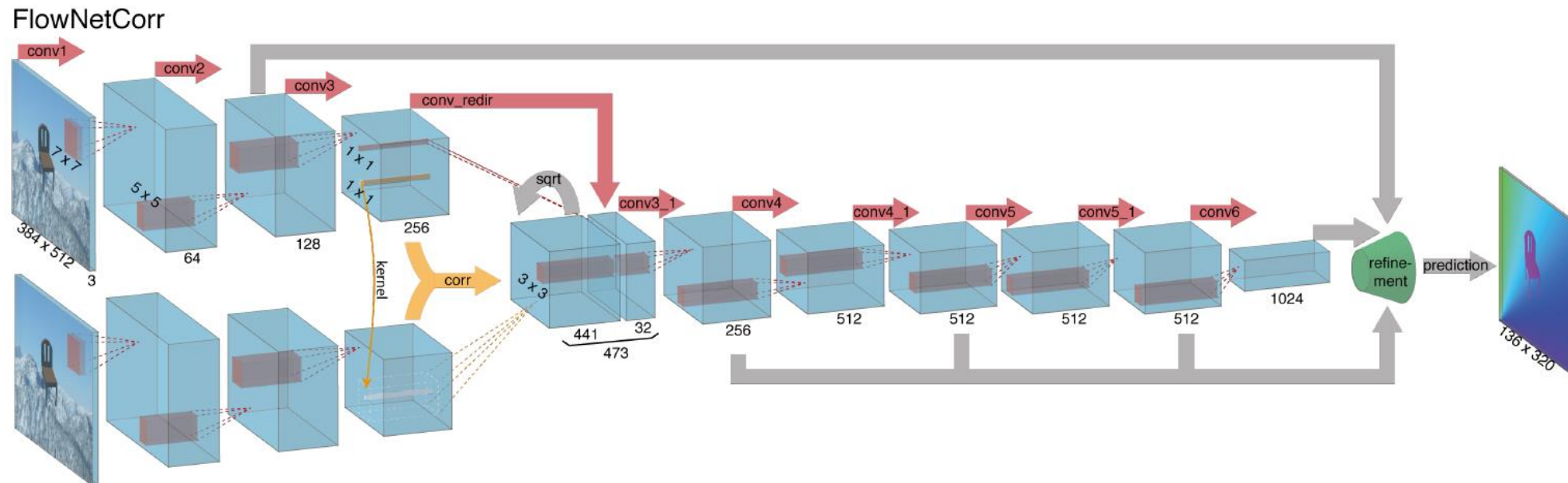
FLOWNET : DÉCODER

- Idée : pour faciliter la tâche, la décomposer en sous-tâches plus simples
 - Petites augmentations successives de résolutions
- Au cours de l'entraînement, le réseau est entraîné à sortir des flux de petites résolutions, puis de moyennes, puis de grandes
 - Se fait en pondérant de manière plus importantes l'erreur de reconstruction des petites résolutions
- Plus la résolution de la sortie est importante, plus on a besoin d'informations précises
 - Envoi des informations des couches adaptées à la résolution pour chaque sous-tâches



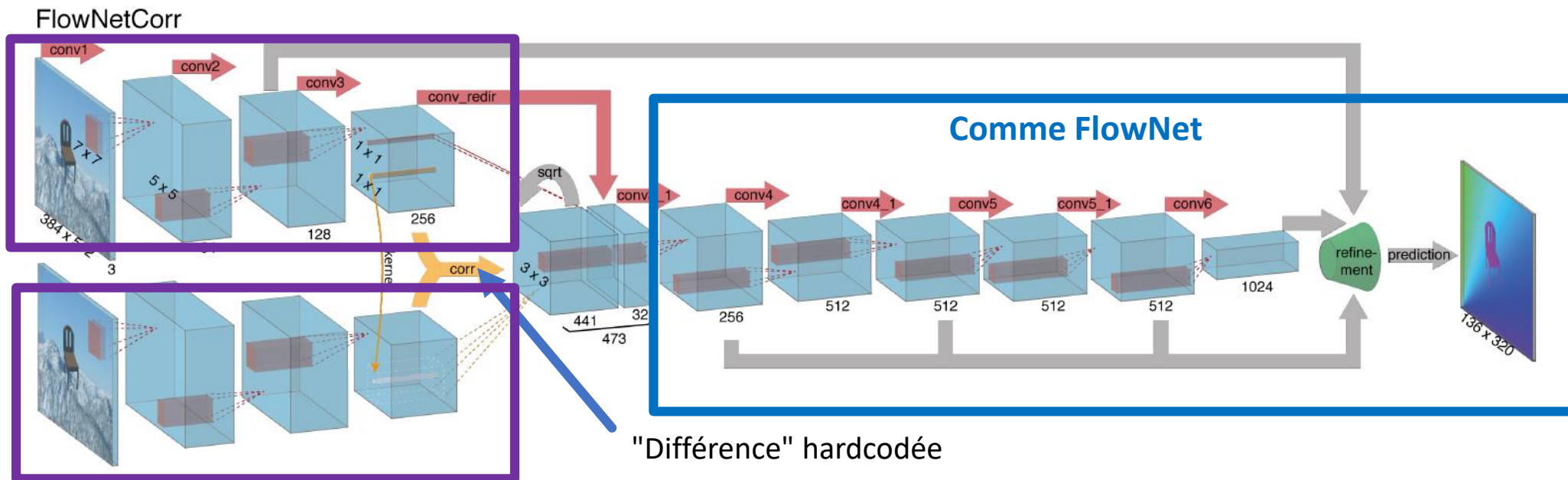
FLOWNETCORR

- ▶ Version améliorée de FLOWNET
 - Facilite la tâche en "hardcodant" le fait qu'il faille comparer la différence entre les deux images
- ▶ La partie superieure de l'encoder et le décodeur sont identiques
- ▶ Extrait des features indépendamment des deux images, avant de fusionner les deux
 - Fusion par "Correlation" hardcodé (non apprise). La corrélation fonctionne par patches locaux.



FLOWNETCORR

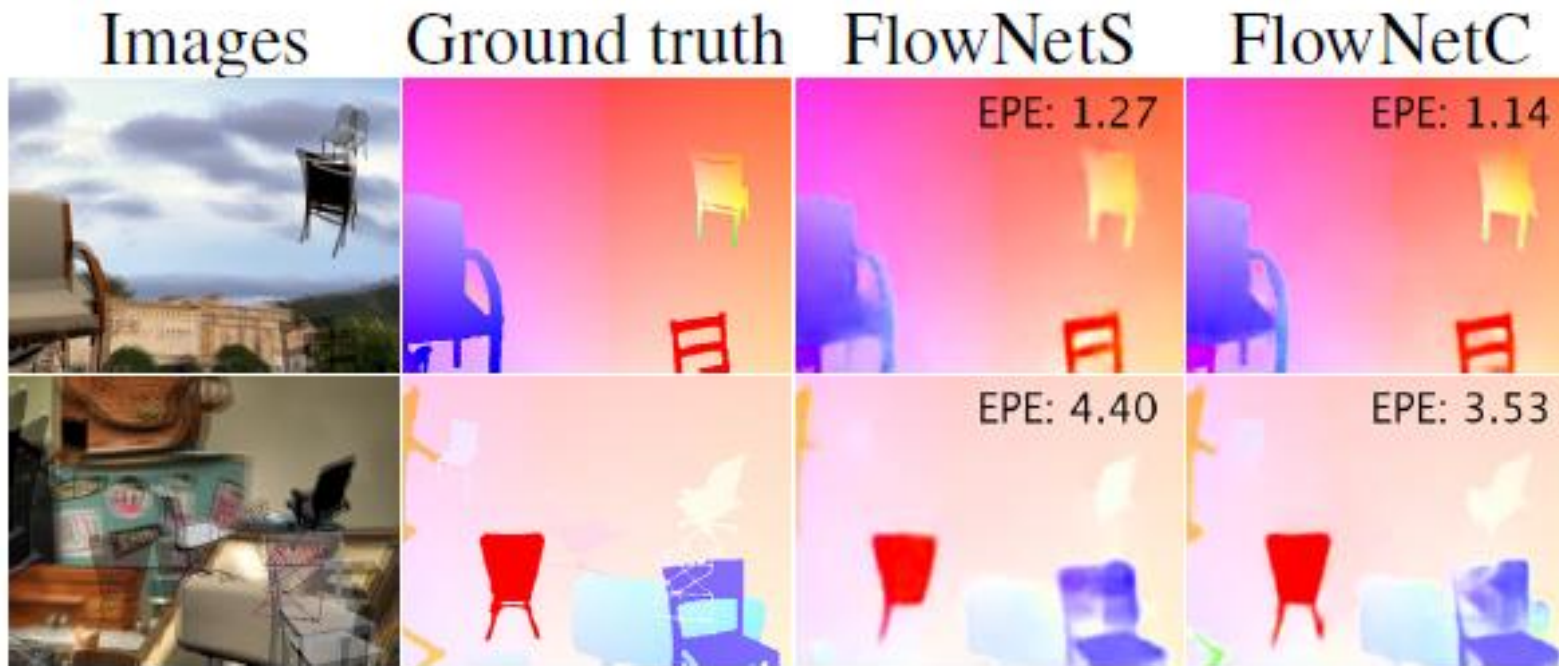
- Version améliorée de FLOWNET
 - Facilite la tâche en "hardcodant" le fait qu'il faille comparer la différence entre les deux images
- La partie supérieure de l'encoder et le décodeur sont identiques
- Extrait des features indépendamment des deux images, avant de fusionner les deux
 - Fusion par "Correlation" hardcodé (non apprise). La corrélation fonctionne par patches locaux.



Extraction de features indépendante sur chaque image

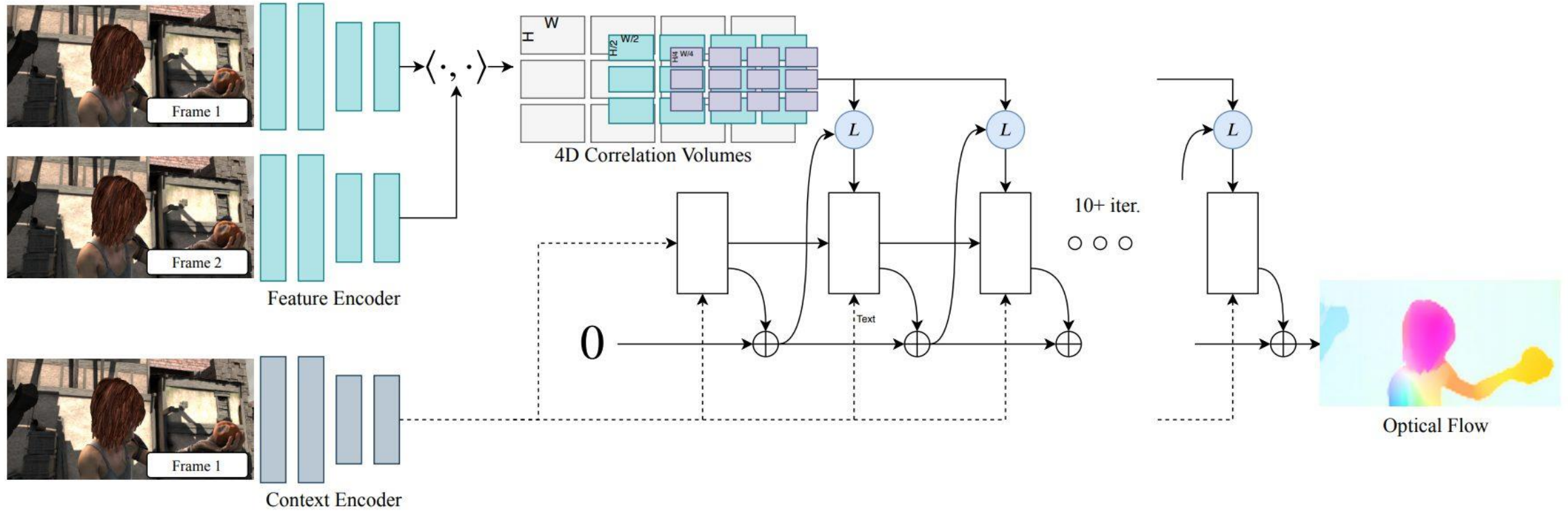
EN EXEMPLE

- Fournit par l'auteur sur le "flying chair dataset"
 - Dataset facile et peu réaliste



RAFT

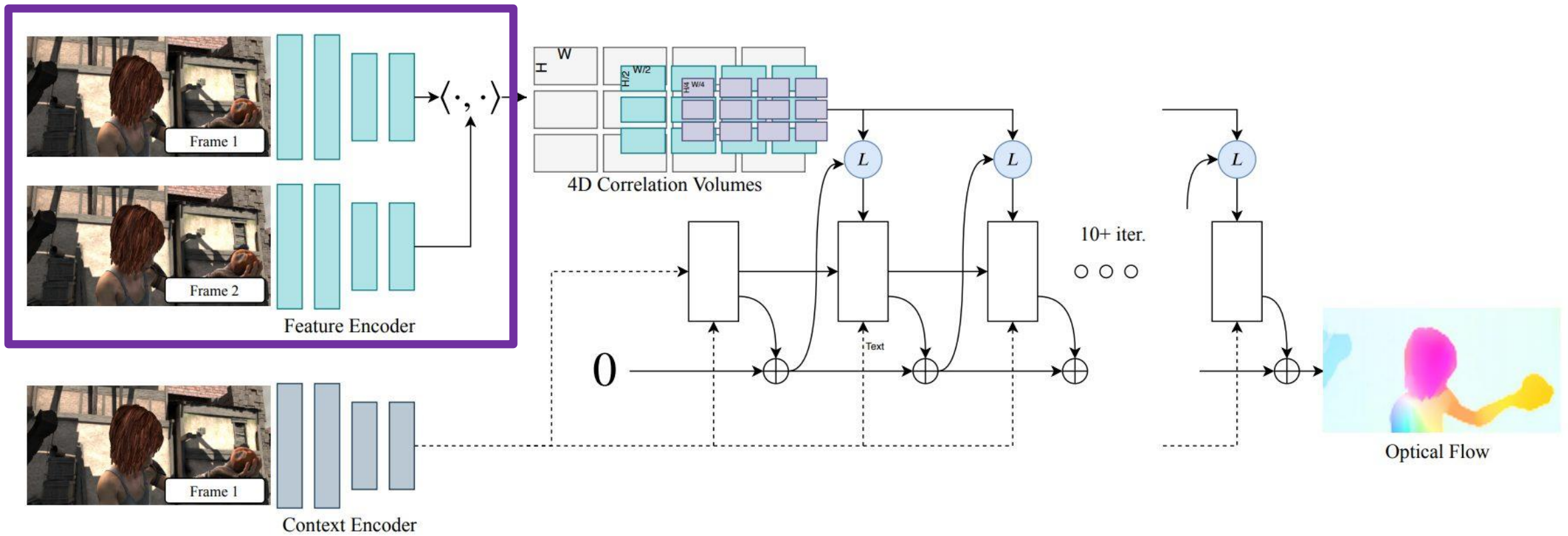
► Raft : Recurrent All-Pairs Field Transforms



RAFT

- Raft : Recurrent All-Pairs Field Transforms

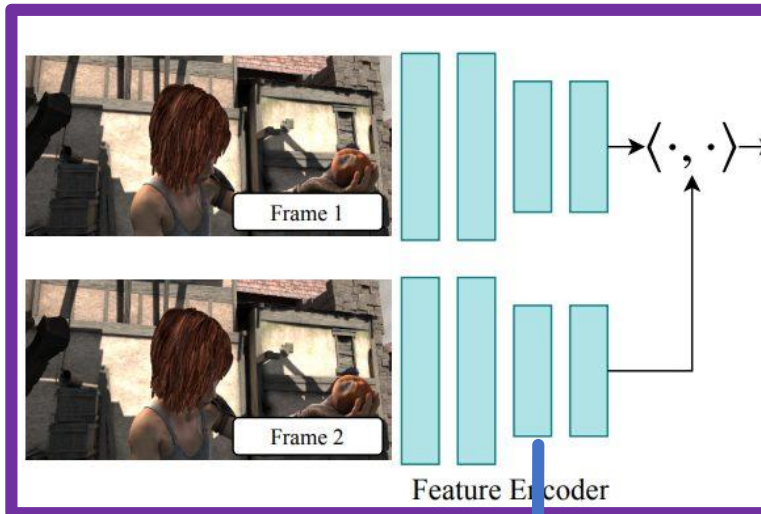
Exactement le début de FlowNetCorr ! (mais avec Resnet)



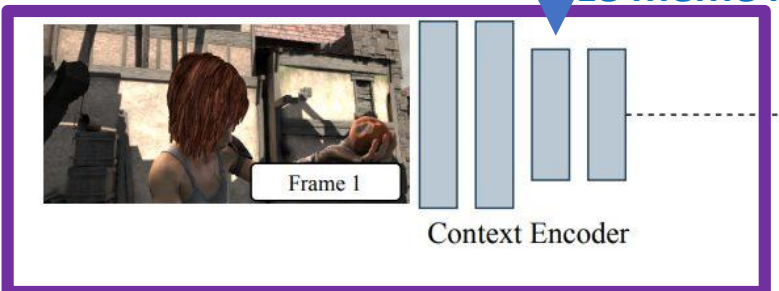
RAFT

- Raft : Recurrent All-Pairs Field Transforms

Exactement le début de FlowNetCorr ! (mais avec Resnet)



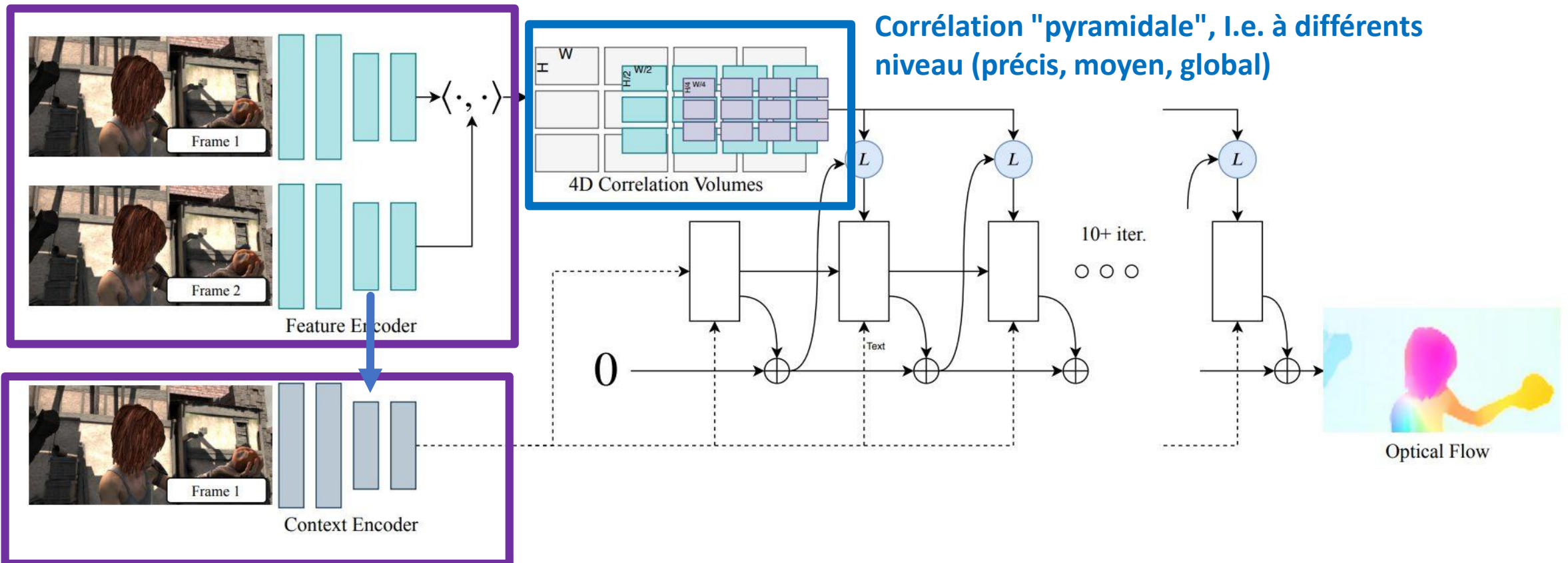
Le même réseau, utilisé pour extraire des features "de contexte"



RAFT

- Raft : Recurrent **All-Pairs** Field Transforms

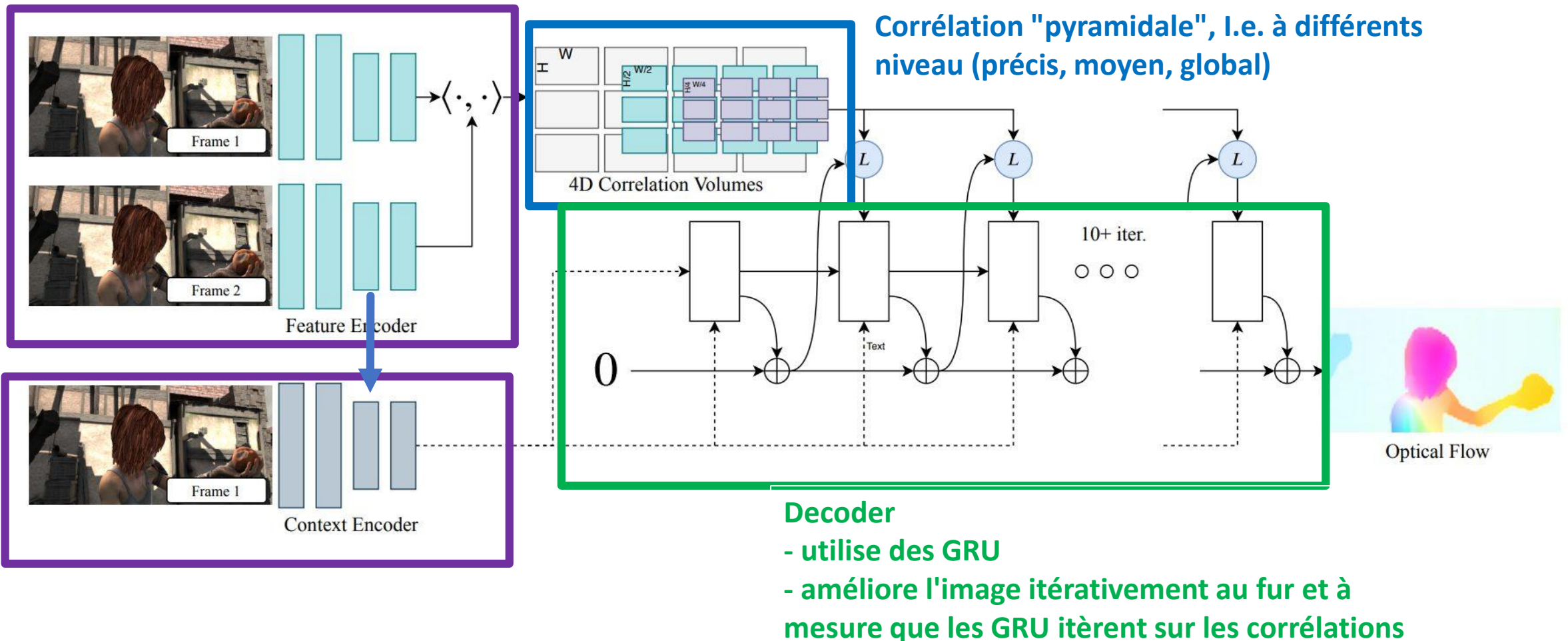
Le début de FlowNetCorr !



RAFT

- Raft : **Recurrent** All-Pairs Field Transforms

Le début de FlowNetCorr !



CONCLUSION

- ▶ Démo
- ▶ Raft marche bien !
 - Si la caméra ne bouge pas, problème de la détection de flux optique.
 - Ne fait pas d'upsampling
- ▶ Raft est lourd ?
 - Des corrélations qui nous font manipuler des matrices 4D.
 - Architecture pensée pour être optimisée
 - Auteurs : 9 fps sur gtx 1080 Ti,
- ▶ Disponible en version "small"
 - Même principe, juste moins de couches et de neurones par couches
 - 20 fps sur 1080 Ti



SCALIAN

A BRIDGE TO PERFORMANCE

f o in t
scalian.com