Drag Your GAN: Iterative Point-based Manipulation on the Generative Image Manifold

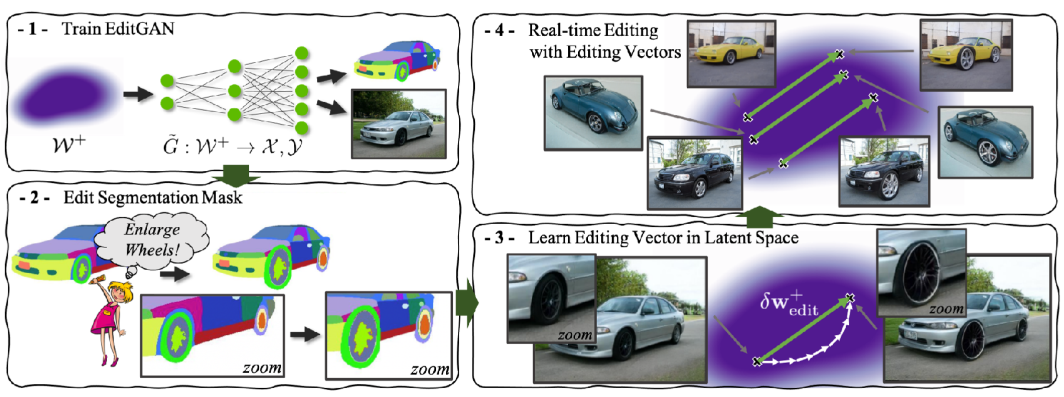
이미지처리팀-류채은

Summary:
Deformation by point-based manipulation performed on the learned image manifold of a GAN, which tends to obey the underlying object structures!

[2. Related Works]

2.1 Generative Models for Interactive Content Creation

- **Unconditional GANs**
  : Do NOT directly enable controllable editing of the generated images (ex. StyleGAN)
  (= 생성된 이미지 통제 불가능!)
- **Conditional GANs**
  : The network receives CONDITIONAL INPUT
  (=condition을 image와 함께 넣어줌!)

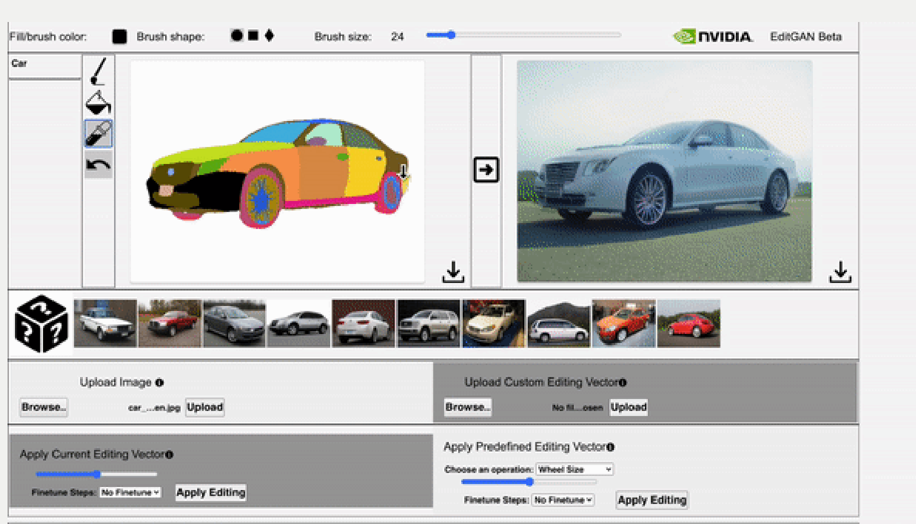


EditGAN Methodology

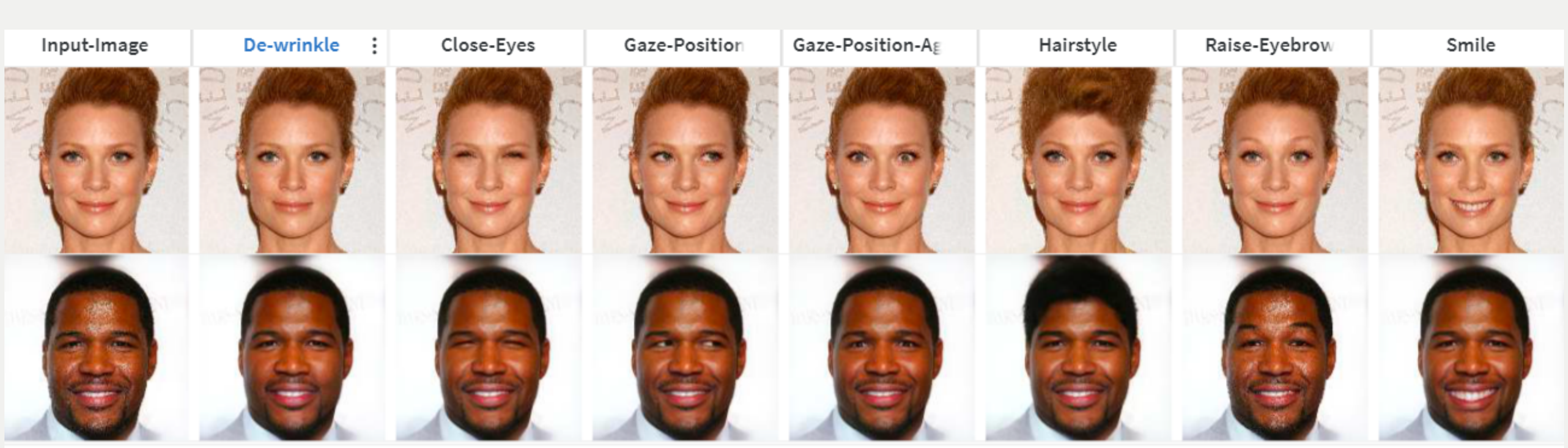
example) EditGAN (https://arxiv.org/pdf/2111.03186.pdf)

Step 1. Learn joint distribution of images and segmentation maps(=conditional input)
Step 2. Compute new images corresponding to the edited segmentation maps

Overview in GIF:



Samples from Experiments of EditGAN



- **Controllability using Unconditionaly GANs**

example) Editing in Style: Uncovering the Local Semantics of GANs (https://openaccess.thecvf.com/content_CVPR_2020/papers/Collins_Editing_in_Style_Uncovering_the_Local_Semantics_of_GANs_CVPR_2020_paper.pdf)

However, it can only achieve the edited image by transferring local semantics between different samples

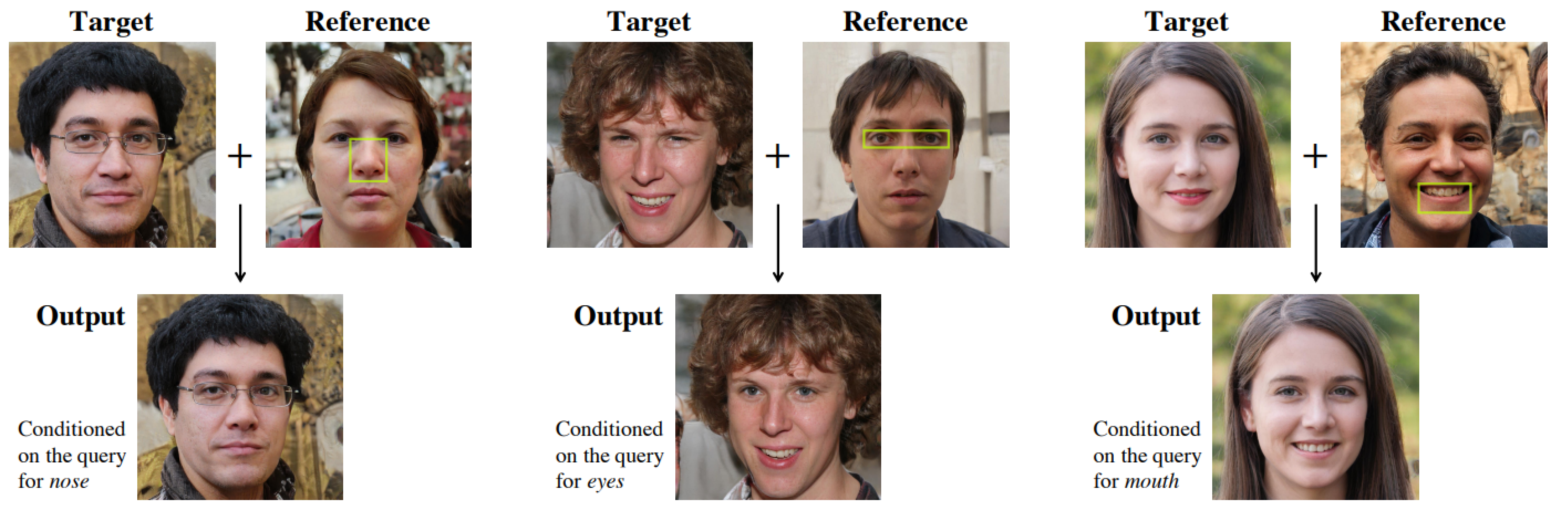


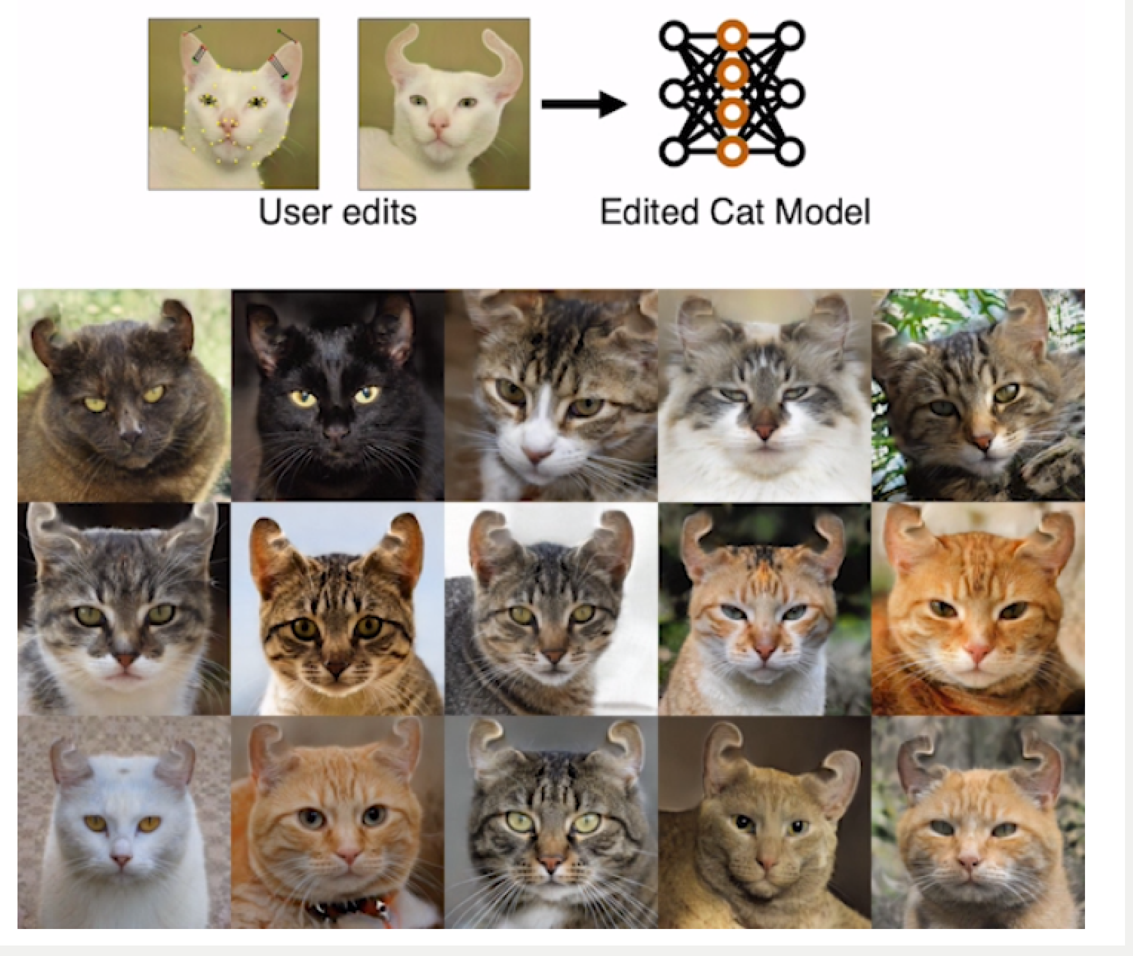
Figure 1: Our method performs local semantic editing on GAN output images, transferring the appearance of a specific object part from a reference image to a target image.

Target: GAN ouput images!

The paper conditioned style-tranfer on local parts of reference image

- **Point-based Editing**

example) GANWarping (https://arxiv.org/pdf/2207.14288.pdf)
(warp 의미: 휨)
Also use point-based editing!
However, does NOT ensure that warps lead to realistic images



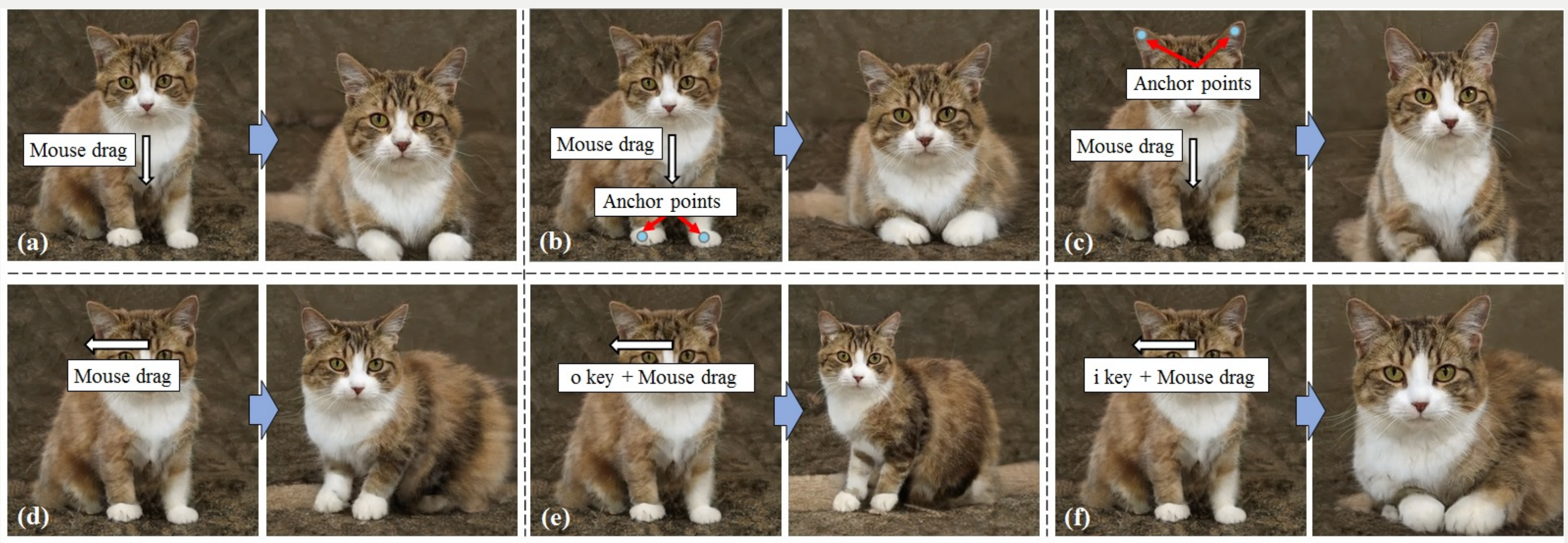
GANWarping demo

example 2) UserControllableLT (https://arxiv.org/pdf/2208.12408.pdf)

: enables point-based editing by transforming latent vectors!

However,

- only supports a single point drag (does NOT handle multiple-point constraints well)
- control is not precise (잘 안된다..)



UserControllableLT demo

sample 1

sample 2

- **3D-aware GANs**

example) Efficient Geometry-aware 3D Generative Adversarial Network (https://arxiv.org/pdf/2112.07945.pdf)

: Generates 3D representations that can be rendered using a physically-based analytic renderer

However, *only limited to global pose*

Figure 1. Our 3D GAN enables synthesis of scenes, producing high-quality, multi-view-consistent renderings and detailed geometry. Our approach trains from a collection of 2D images without target-specific shape priors, ground truth 3D scans, or multi-view supervision. Please see the accompanying video for more results.

sample 1 (global view 조절 정도만 가능하다!)

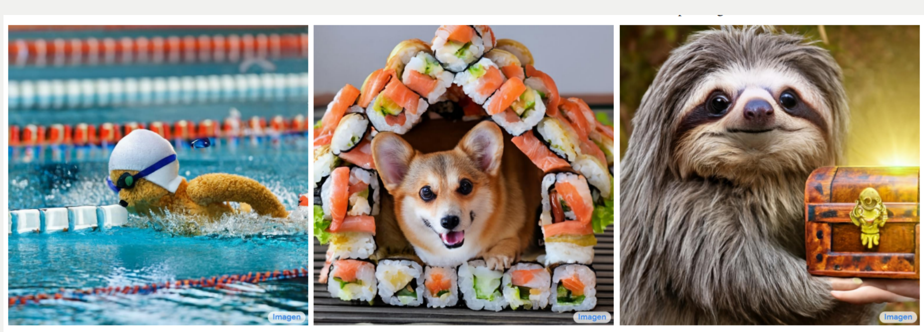
- **Diffusion Models**

공통적인 한계점: 느리다! (GAN이 훨씬 efficient하다)

example) Photorealistic Text-to-Image Diffusion Models
with Deep Language Understanding (https://imagen.research.google/paper.pdf)

: Recent models have shown expressive image synthesis conditioned on text inputs.

However, fine-grained control over the spatial attributes of images is NOT supported with natural language!



Teddy bears swimming at the Olympics 400m Butterfly event. A cute corgi lives in a house made out of sushi. A cute sloth holding a small treasure chest. A bright golden glow is coming from the chest.

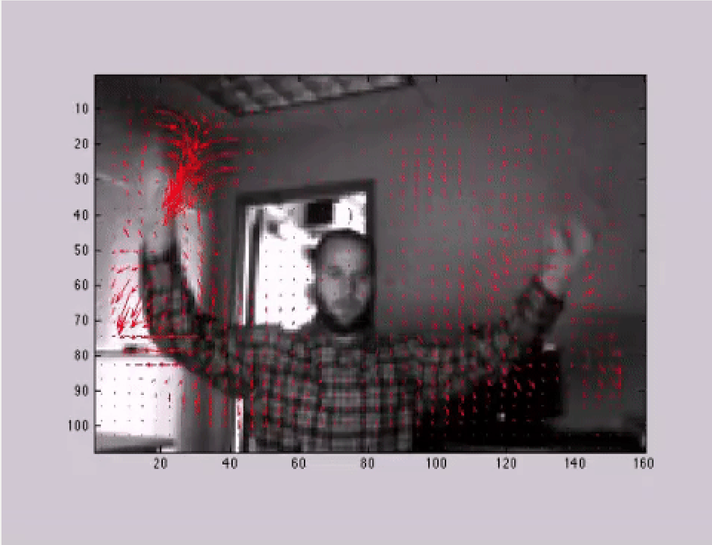
samples from Imagen

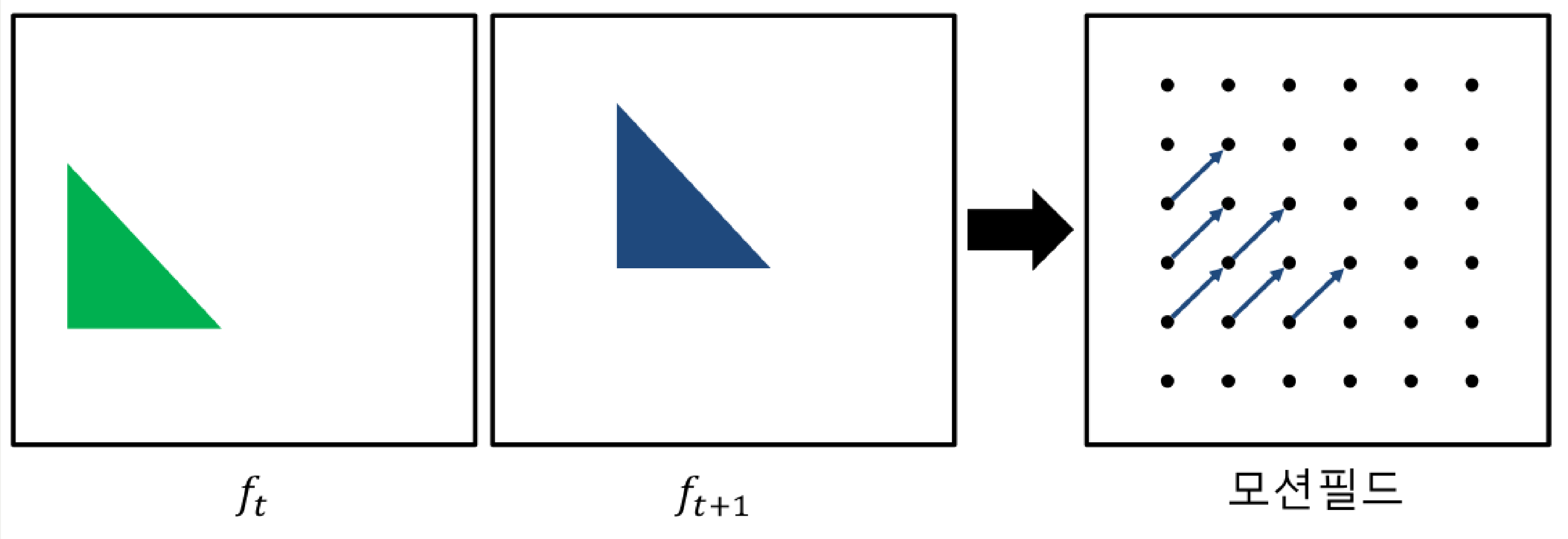
2.2 Point Tracking

Prerequisites: Optical Flow Estimation이란?

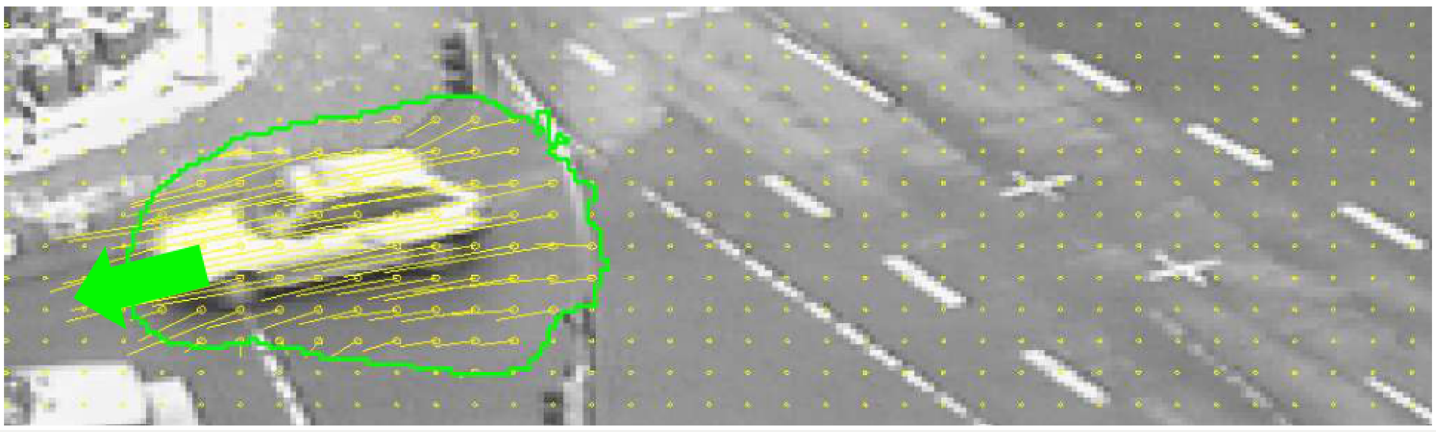
Optical Flow

: The pattern of apparent motion of image objects between two consecutive frames caused by the movement of object or camera
(=Optical Flow란 video 에서 frame 별로 object가 어떻게 움직이는 지 관찰하는 task)





모션필드 (=optical field): 움직임이 발생한 모든 점의 모션 벡터로 얻어낸 2차원 모션맵



optical flow estimation을 통해 물체의 움직임을 구분해낼 수 있다!

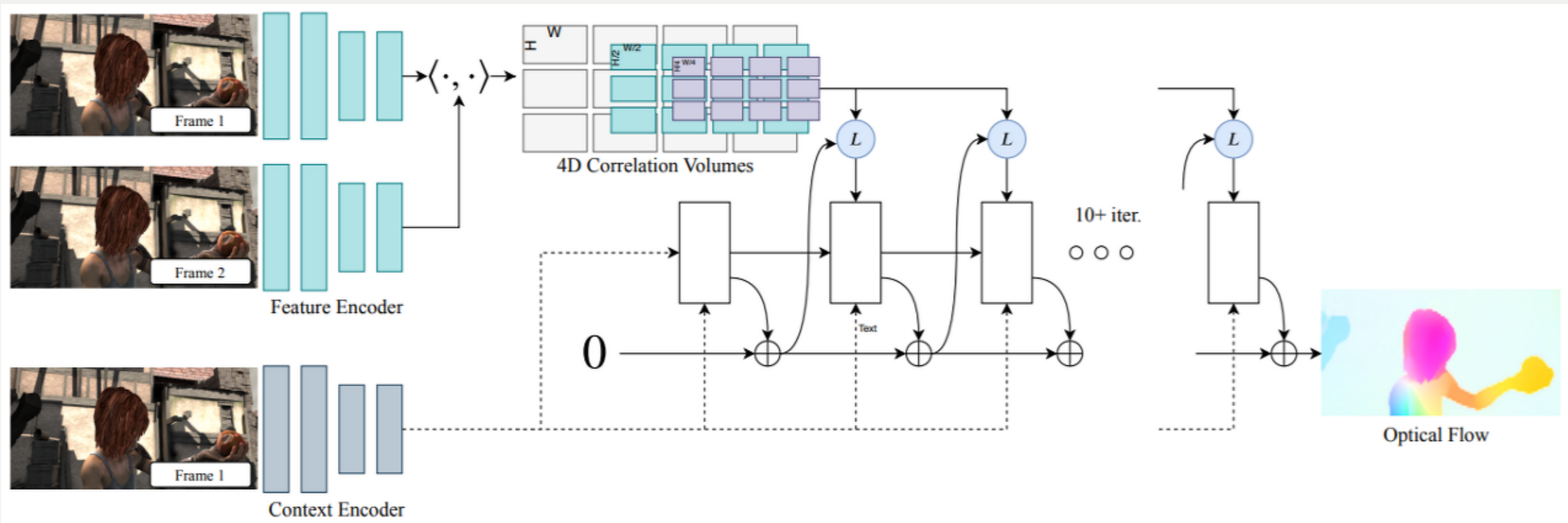
Clarification) motion이 생김에 따라 발생하는 **명암변화**를 계산함으로서 motion field 역할을 하는 map을 생성하는게 optical flow! (ref: https://searching-fundamental.tistory.com/16)

많이 사용되는 DL-based Point Tracking Approach example) RAFT, Recurrent All-Pairs Field Transforms for Optical Flow (https://arxiv.org/pdf/2003.12039.pdf)

: Optical Flow와 Transformer를 결합하면서 SOTA를 달성

1. Image간 feature을 extract한다.
2. 두 쌍의 correlation을 구한다.
3. flow field를 iterative하게 refine하여 update한다.

(ref: https://searching-fundamental.tistory.com/38)



RAFT algorithm overview

하지만, RAFT보다 DragYourGAN의 point-tracking 성능이 더 좋다!