2023 여름 세미나

2023.08.18

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented By
이준호

Outline

- Intro
  - What is inpainting & outpainting
  - What is different image outpainting & video outpainting
- Inpainting for Video outpainting
  - Onion peel network & Copy and Paste network
  - Flow-based method
    - Optical flow
- 2020 ECCV Flow based Video inpainting paper
  - Flow-edge Guided Video Completion
- 2021 ICLR Flow based Video inpainting paper
  - Large scale image completion via Co Modulated Generative Adversarial Networks
- 2022 CVPR Video outpainting paper
  - Complete and temporally consistent video outpainting
- Conclusion & Appendix

Intro

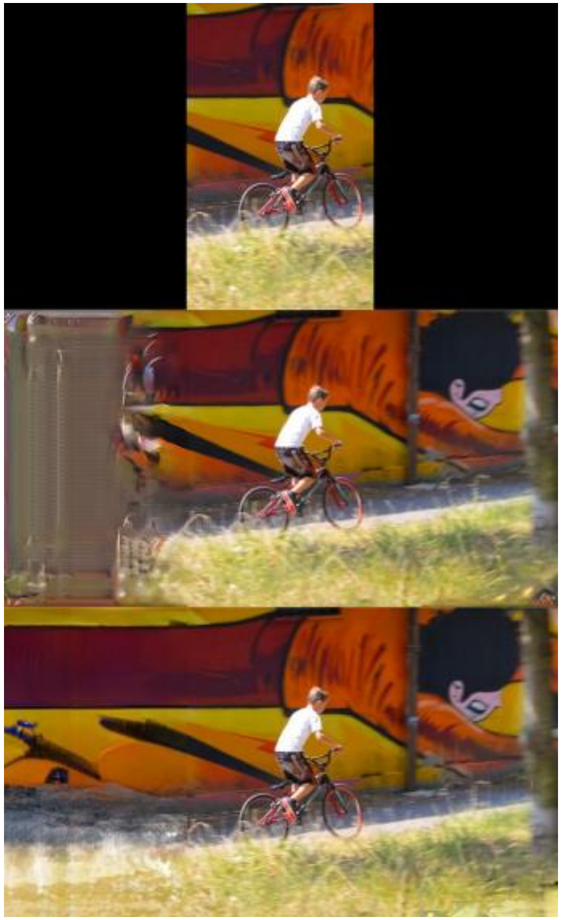
- What is inpainting
  - 빈 영역이 input image/video 내에 위치
  - 모든 방향에 대한 정보가 제공
  - 생성되는 영역이 input image에 비해 작음

(a) Original images (b) Masked images (c) Inpainted images

- What is outpainting
  - 빈 영역이 input image/video의 한쪽 옆이나 주변에 있음
  - 빈 영역의 한쪽에만 알려진 픽셀이 있음
  - 생성되는 영역이 input image의 몇 배

Intro

- What is different image outpainting & video outpainting
  - Image outpainting
    - 정적인 image에 대한 작업
  - Video outpainting
    - Video의 frame마다 image의 외부 영역을 생성하는 작업
    - 각 frame마다 움직이는 object와 background가 변화
      - Time 축을 갖음
      - 이를 일관성 있게 처리하는 작업이 중요



2022 CVPR Video outpainting paper

- Inpainting for Video Outpainting
  - 첫번째 video inpainting method : Onion-Peel network
    - 각 frame마다 이전 frame의 예측 결과를 활용하여 blank부분을 예측하고 이 과정을 반복해서 영상의 연속성을 유지
  - 두번째 video inpainting method : Copy and Paste network
    - 현재 frame에서 blank부분을 주변 pixel의 정보를 활용하여 채운 뒤 다음 frame에 위 결과물을 복사하고 붙여 넣음
    - 이 과정을 반복해서 time consistency를 유지
  - 두 논문은 이전 frame을 이용해 현재 frame을 보완하는 공통점을 갖음
  - 그러나 foreground가 frame 밖으로 벗어나면 temporal artifacts가 발생

2022 CVPR Video outpainting paper

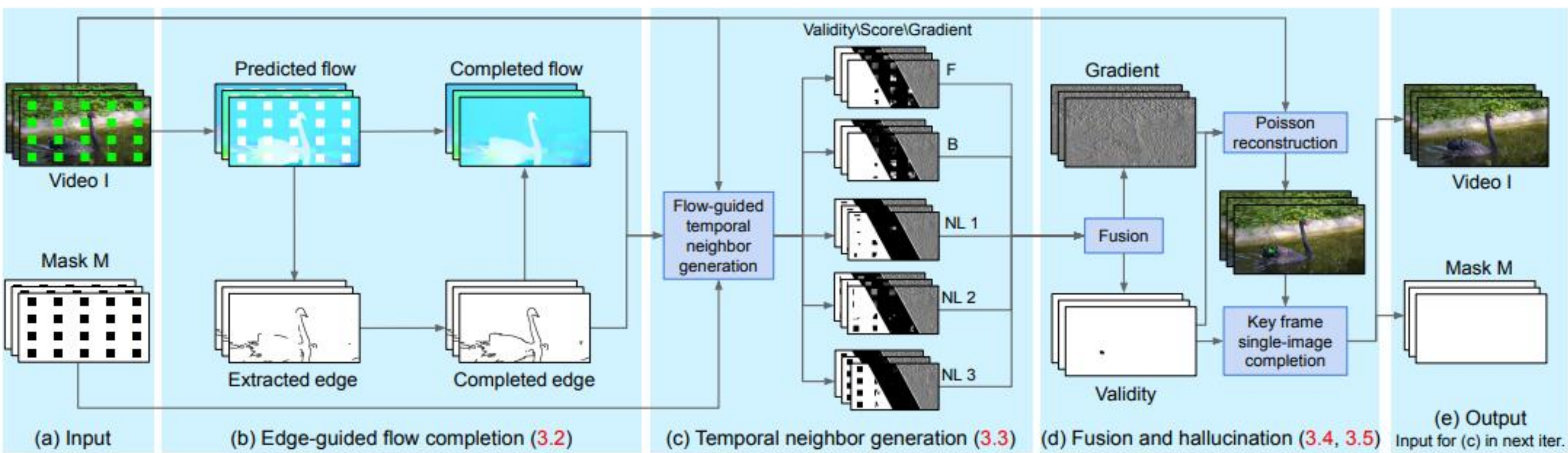
- Inpainting for Video Outpainting
  - Flow-based method
    - Time consistency를 유지하려면 flow-based method가 가장 적합하다 판단
    - Optical flow를 method로 채택
  - Optical flow란
    - Optical field를 구하기 위해 이전 frame과 현재 frame의 차이를 이용하고 각 픽셀의 이동을 계산하여 object의 움직임을 구별하는 방법
  - Optical flow의 2가지 가정
    - Color/brightness constancy : 어떤 픽셀과 주변 픽셀의 color/brightness는 같음
    - Small motion : frame간 움직임이 작아 픽셀 점은 멀리 움직이지 않음

2022 CVPR Video outpainting paper

- Flow-based method
  - Flow-edge Guided Video Completion_ECCV 2020
    - 요약
      - Motion edge를 extract하고 complete
      - Color와 flow를 합성하고 flow를 따라 color를 propagate
      - Main propose는 object boundaries를 따라 sharp한 flow edge를 propagate
    - Main Contribution
      - Flow edge : sharp하게 flow edge를 얻어 smoot한 flow edge를 얻음
      - Non-local flow : Non-local flow를 활용하여 transitive flow를 처리
      - Seamless blending : visible seams를 피하기 위해 gradient domain에서 작업

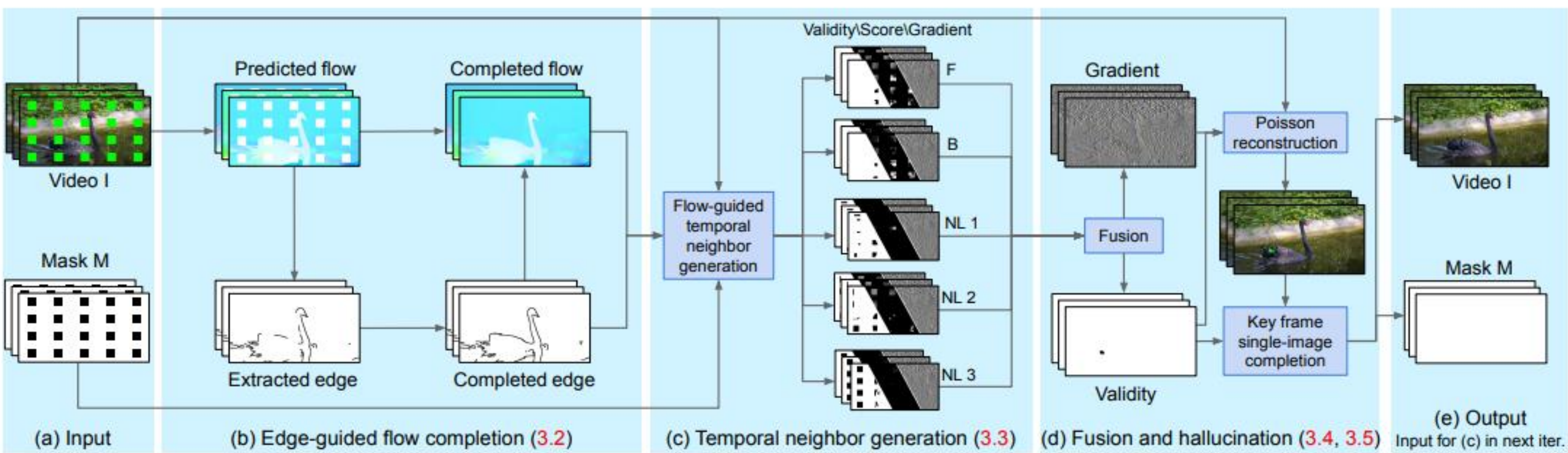
2022 CVPR Video outpainting paper

- Flow-based method
  - Flow-edge Guided Video Completion_ECCV 2020
    - Input : 합성해야 할 color video와 binary mask
    - Edge-guided flow completion : 인접한 frame, 인접하지 않은 frame 사이의 forward, backward 간의 flow를 계산하고, flow edge를 extract 및 completion하여 smooth한 flow completion을 유도
    - Temporal neighbor generation : missing pixel에 대한 후보 pixel 집합을 계산하기 위해 flow trajectories를 만듦



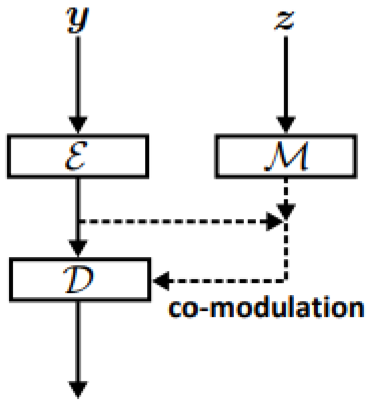
2022 CVPR Video outpainting paper

- Flow-based method
  - Flow-edge Guided Video Completion_ECCV 2020
    - Fusion and hallucination : confidence-weighted average를 사용하여 각 missing pixel에 대한 gradient domain 후보를 통합, 그 중 pixel이 가장 많이 빠진 frame을 선택하고 inpainting
    - Output : missing pixel이 없을 때 까지 위 과정의 결과물을 다음 iteration으로 전달



2022 CVPR Video outpainting paper

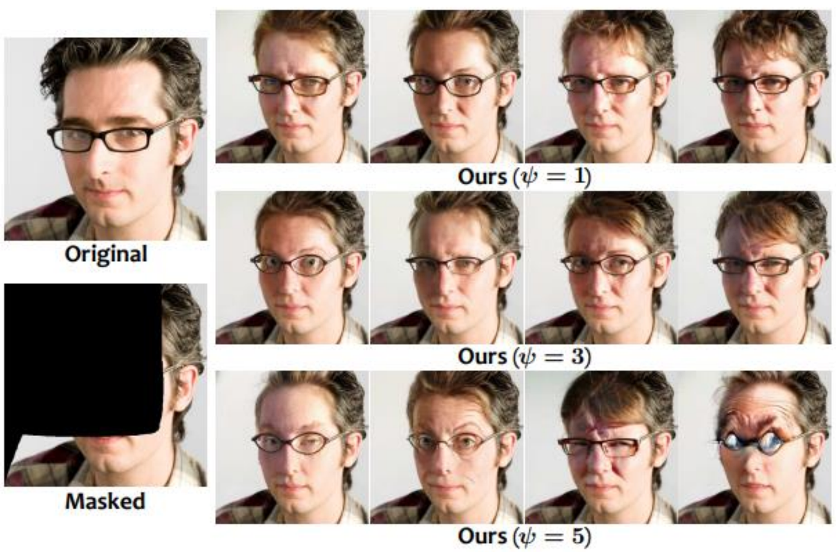
- Image complete method
  - Large scale image completion via Co Modulated Generative Adversarial Networks 2021_ICLR
    - 기존의 unconditional GAN은 latent vector에 의해 생성된 learned style을 활용
    - 위 방법론을 image-conditional GAN에도 적용하려 했지만 stochasticity가 부족하여 limited conditional information이 제공되는 환경에서 적용하기엔 어려움
      - 즉 mask(blank) 된 영역이 크다면 생성하기 어렵다는 뜻
    - 위 문제를 해결하기 위해 Co-modulation을 제안



$$s = \mathcal{A}(\mathcal{E}(y), \mathcal{M}(z))$$

2022 CVPR Video outpainting paper

- Flow-based method
  - Large scale image completion via Co Modulated Generative Adversarial Networks 2021_ICLR
    - Style space에서 style vector가 linearly correlated하다는 것을 가정했을 때 성능 개선
    - Co modulated GAN은 discriminator loss와 함께 훈련
    - L1 term처럼 직접적인 guide가 필요하지 않아 stochastic generative capability를 활용

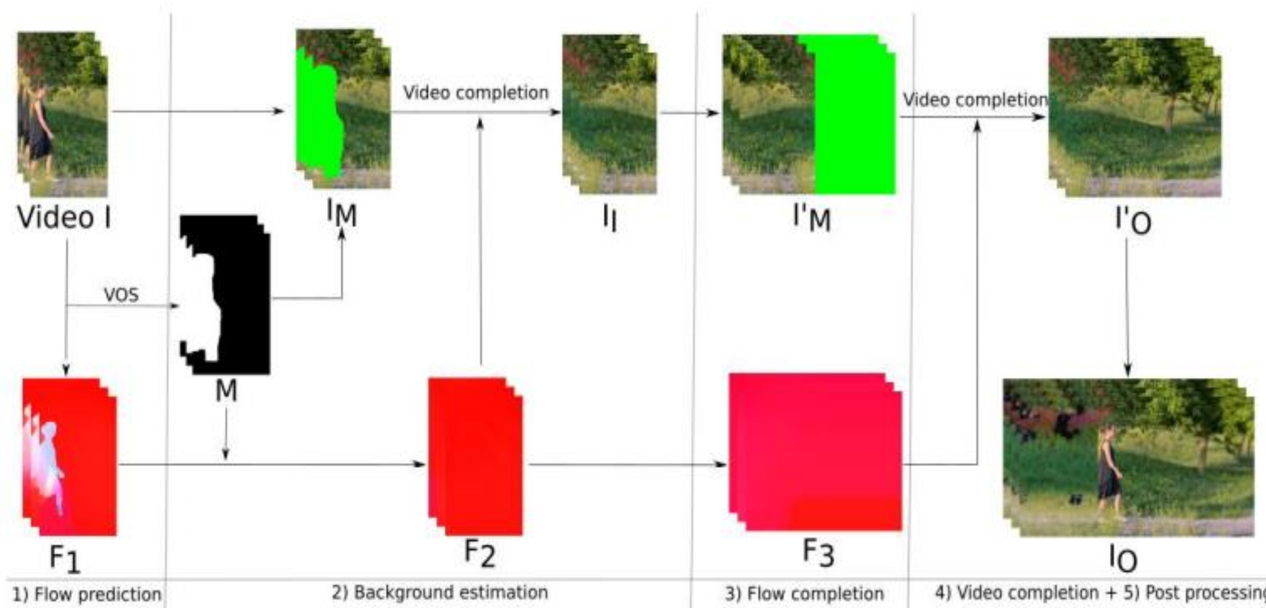


2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting CVPR_2022(Workshop)
  - 요약
    - Video inpainting을 outpainting에 적용할 때의 단점을 확인
    - Image shifting을 사용하여 image outpainting을 개선하는 방법, outpainting method를 설명
  - Main Contribution
    - 원본 video content를 손상시키지 않고 시각적으로 만족스러운 temporally consistent한 completions를 제공
    - Optical flow를 사용하여 인접한 frame 간의 정보를 전달하여 temporally consistent를 형성
    - Background estimation을 초기에 진행하여 temporal artifacts를 줄임

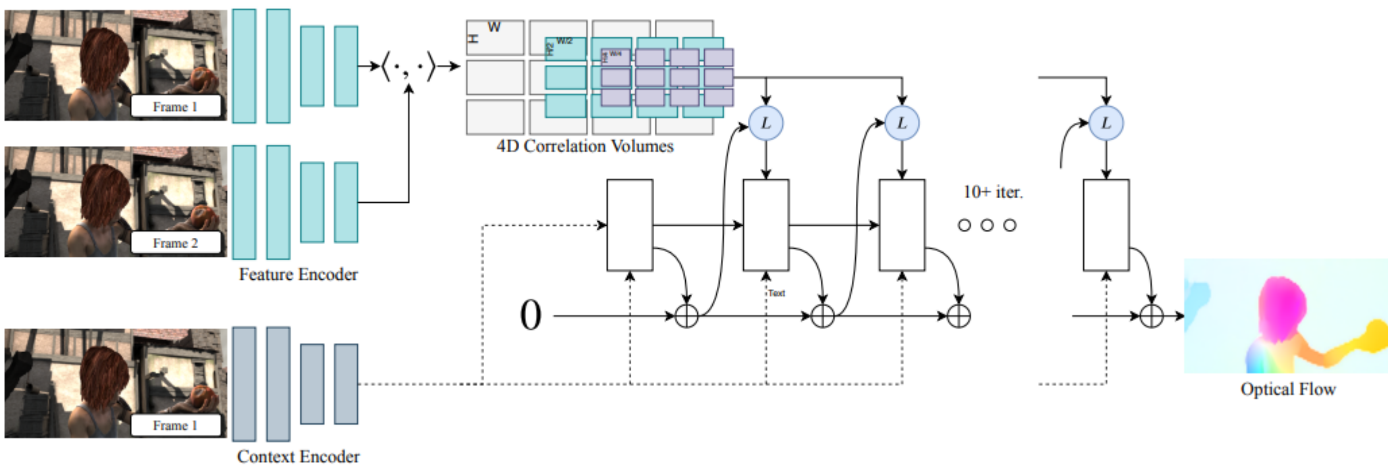
2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Overview of method
    - Flow estimation
    - Background estimation
    - Flow completion
    - Video completion
    - Post-processing



2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Video Completion (RAFT ECCV_2020)
    - Color propagation을 사용하여 optical flow를 기반으로 빈 영역의 일부를 완성시킴
    - Optical flow를 추정하기 위해 현재 flow estimation method S.O.T.A인 RAFT를 사용함
    - Optical flow를 완성하기 위해 masking된 영역 내의 gradient를 최소화 시킴
      - 움직이는 object를 제거하거나 background의 일부를 제거했을 때 더 부드럽게 생성된다고 함
    - 그러나 위 method는 움직이는 foreground objects가 video outpainting에 적용될 때 temporal artifacts가 발생

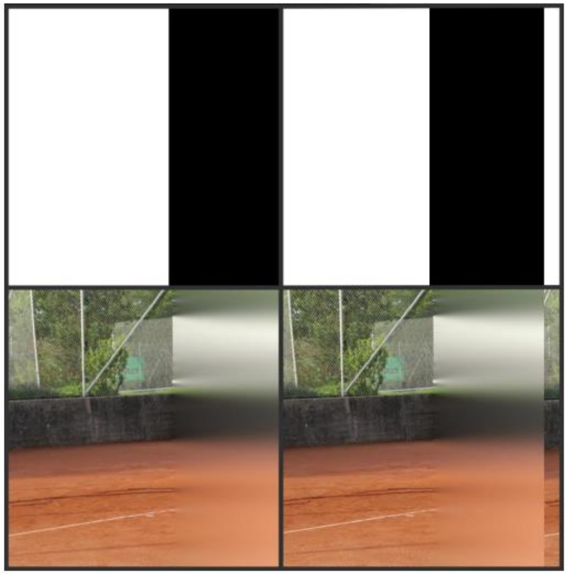


2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Background Estimation (Flow-edge Guided Video Completion를 사용)
    - Frame의 edge를 따라 물체가 움직일 때, object가 not visible할 수 있음
    - Foreground motion의 complete outline이 없기 때문에 foreground, background motion이 혼동
      - Optical flow를 estimation하는데 문제가 발생
    - 그래서 초기에 background estimation을 진행
    - Background estimation을 하기 위해 우선 foreground mask가 필요하기 때문에 VOS(Video Object Segmentation) method를 채택
      - VOS란 foreground object를 background와 separate하는 binary labeling problem
    - Background에 foreground mask를 이용해서 masking하고 inpainting method를 사용하여 채움

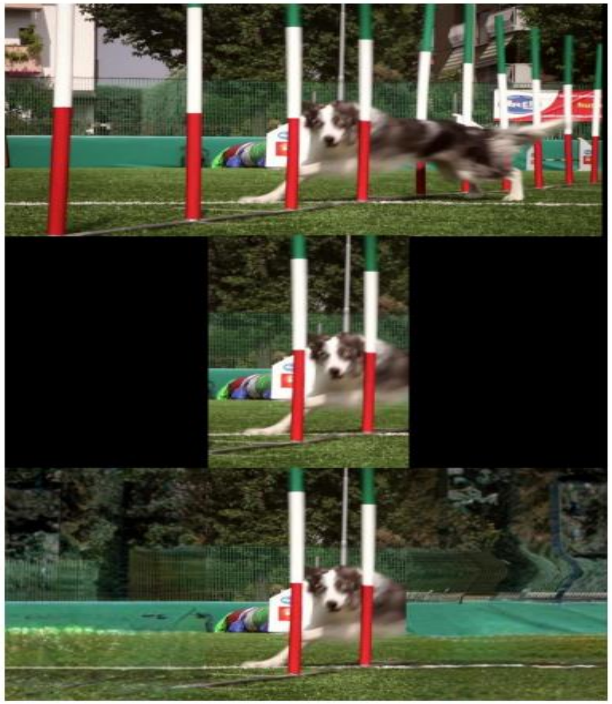
2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Image Completion
    - Flow edge guided network를 image outpainting network로 사용하려 했으나 large-scale image completion network를 사용
    - Image content를 shift하여 blank 영역의 바깥쪽에 추가 정보를 넣음
      - 실험적으로 더 smooth한 결과를 생성하는 것을 입증
    - 가장 오른쪽의 known pixel을 반사(mirror)해서 추가 정보 제공



2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Post-Processing
    - Frame의 outside information을 predict하는 것은 불가능하다 주장
    - Complete된 영역을 blurry하게 만들어서 original video와 합침
    - 본 논문에선 평가 과정 중 blurring 과정을 포함하지 않음



서강대학교
SOGANG UNIVERSITY

VDS LAB

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - DATASET
    - Image completion network : Places scene recognition dataset (2.5 million images)
    - DAVIS(Densely Annotation Video Segmentation dataset)
      - Two resolution : 480p, 1080p
      - 50 video seqeunces with 3455 annotated frames at pixel level
    - Youtube-VOS(Video Object Segmentation)
      - Large-scale benchmark (multiple VOS task, semi-supervised VOS, VOS)
      - More than 4000 high-resolution Youtube-video & 340 minutes video
    - 본 논문에서는 video frame의 왼, 오른쪽을 crop한 후 input으로 사용
    - Foreground의 annotation을 사용하지 않음

(a) *A person wearing a white shirt with white helmet riding a bike.*

(b) *A laying cat gets up and jumps towards the camera.*

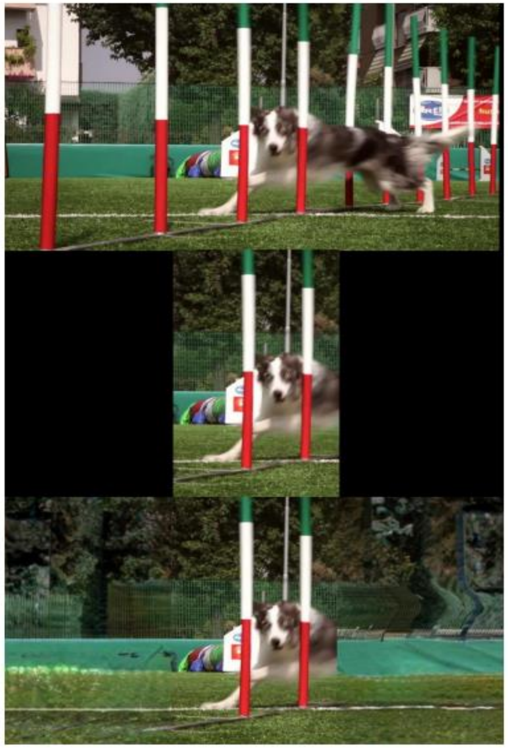
2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Evaluation metric
    - MSE(Mean Squared Error)
    - PSNR(Peak Signal To Noise Ratio)
    - SSIM(Structural Similarity Index Measure)
    - LPIPS(Learned Perceptual Image Patch Similarity)
    - FVD(Frechet Video Distance)

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Portrait To Landscape Conversion
    - Portrait (9:16) to Landscape (16:9)
      - 원본 video의 왼, 오른쪽 edge를 제거하여 input으로 사용
    - 비교 대상 방법
      - 1. Flow-edge Guided Video Completion
      - 2. Ours without image shifting
      - 3. Ours with image shifting
      - 4. Ours with both image shifting and post processing (표 X)

DAVIS dataset [22]	MSE↓	PSNR ↑	SSIM ↑	LPIPS ↓	FVD ↓
Standard	11293,18	7,95	0,330	0,5397	2009,12
Gao *et al.* [7]	1724,97	16,18	0,560	0,3049	1414,86
Video outpainting (ours)	1654,59	16,82	0,596	0,2635	1244,77
Video outpainting+image shift (ours)	**1513.49**	**17.33**	**0.600**	**0.2530**	**1099,11**
YouTube-VOS [27]	**MSE↓**	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**	**FVD ↓**
Standard	11271,97	8,177	0,354	0,470	2220,93
Gao *et al.* [7]	3008,74	14,37	0,500	0,385	1848,07
Video outpainting (ours)	2702,43	14,46	0,509	0,338	1642,46
Video outpainting+image shift (ours)	**2604,17**	**14,76**	**0,518**	**0,320**	**1374,85**



2022 CVPR Video outpainting paper

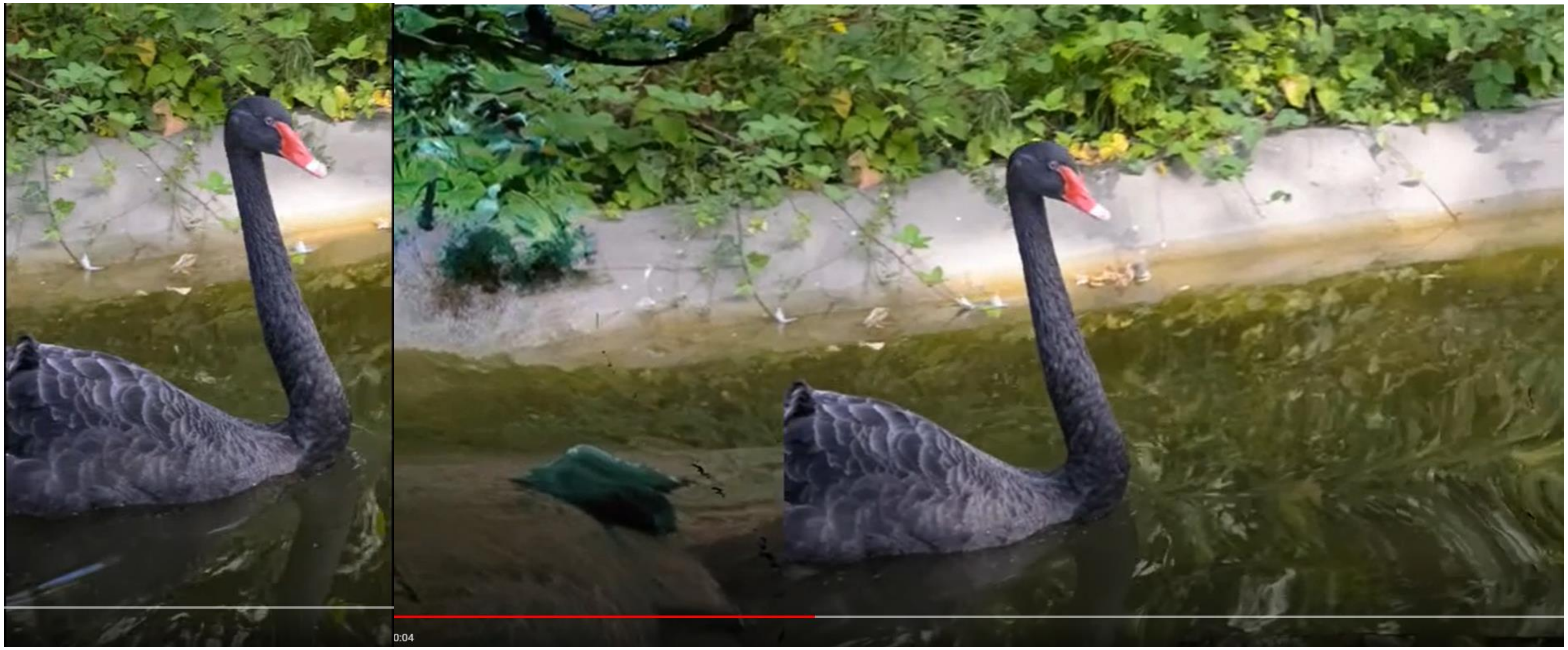
- Complete and Temporally consistent video outpainting
  - Portrait To Landscape Conversion
    - Flow-edge Guided Video Completion

서강대학교 SOGANG UNIVERSITY

VDS LAB

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Portrait To Landscape Conversion
    - Ours without image shifting



SOGANG UNIVERSITY 서강대학교

VDS LAB

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Portrait To Landscape Conversion
    - Ours with image shifting

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Portrait To Landscape Conversion
    - Ours with both image shifting and post processing

서강대학교 SOGANG UNIVERSITY

VDS LAB

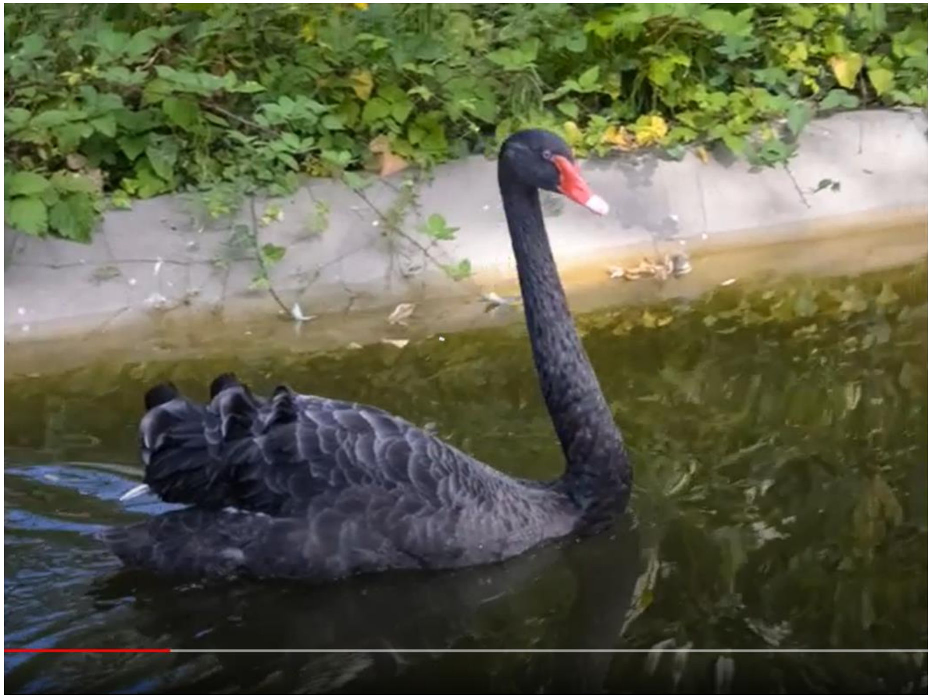
2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Landscape To Ultrawide Conversion
    - Landscape (16:9) to Ultrawide (21:9)
      - DAVIS dataset으로 실험
    - 비교 대상 방법
      - 1. Flow-edge Guided Video Completion
      - 2. Ours without image shifting
      - 3. Ours with image shifting
      - 4. Ours with both image shifting and post processing (표 X)

Method	MSE↓	PSNR ↑	SSIM ↑	LPIPS ↓	FVD ↓
Standaard	11657,35	7,80	0,329	0,546	346,62
Gao et al. [7]	301,33	23,00	0,809	0,074	254,55
Video outpainting (ours)	277,60	23,82	0,852	0,065	224,77
Video outpainting+image shift (ours)	**239,18**	**24,34**	**0,890**	**0,062**	**207,26**

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Landscape To Ultrawide Conversion
    - Original



2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Landscape To Ultrawide Conversion
    - Flow-edge Guided Video Completion

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Landscape To Ultrawide Conversion
    - Ours without image shifting

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Landscape To Ultrawide Conversion
    - Ours with image shifting

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Landscape To Ultrawide Conversion
    - Ours with both image shifting and post processing

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Portrait To Landscape Conversion

DAVIS dataset [22]	MSE↓	PSNR ↑	SSIM ↑	LPIPS ↓	FVD ↓
Standard	11293,18	7,95	0,330	0,5397	2009,12
Gao *et al.* [7]	1724,97	16,18	0,560	0,3049	1414,86
Video outpainting (ours)	1654,59	16,82	0,596	0,2635	1244,77
Video outpainting+image shift (ours)	**1513.49**	**17.33**	**0.600**	**0.2530**	**1099,11**
YouTube-VOS [27]	**MSE↓**	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**	**FVD ↓**
Standard	11271,97	8,177	0,354	0,470	2220,93
Gao *et al.* [7]	3008,74	14,37	0,500	0,385	1848,07
Video outpainting (ours)	2702,43	14,46	0,509	0,338	1642,46
Video outpainting+image shift (ours)	**2604,17**	**14,76**	**0,518**	**0,320**	**1374,85**

  - Landscape To Ultrawide Conversion

Method	MSE↓	PSNR ↑	SSIM ↑	LPIPS ↓	FVD ↓
Standaard	11657,35	7,80	0,329	0,546	346,62
Gao et al. [7]	301,33	23,00	0,809	0,074	254,55
Video outpainting (ours)	277,60	23,82	0,852	0,065	224,77
Video outpainting+image shift (ours)	**239,18**	**24,34**	**0,890**	**0,062**	**207,26**

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - Limitations
    - Foreground object를 outpainting area에서 complete하지 않기로 결정
      - Frame edge 근처의 moving object가 artifacts를 야기함
    - Background estimation을 초기에 했을 때 생기는 단점 2가지
      - Object가 close-up 됐을 때 background의 정보가 매우 부족함
      - Background만 complete할 때 foreground object가 complete area에서 사라짐
    - Video에 빠르거나 복잡한 camera motion이 포함되어있는 경우
      - Optical flow 기반으로 한 complete이 부자연스러움

2022 CVPR Video outpainting paper

- Conclusion
  - 본 논문의 최종 목표는 원본 video의 outpainting area를 blurr하게 처리하는 것
    - 이 목표를 전제로 진행한 연구이기 때문에 연구 방향의 제약이 많다고 생각
  - Video outpainting 분야는 크게 3가지 framework를 따름
    - Temporal consistency를 어떻게 유지할 것인가?
    - Foreground object를 어떻게 처리할 것인가?
    - Outpainting method를 어떻게 할 것인가?
  - Framework의 접근 방법에 따라 연구 novelty가 달라진다고 생각함
    - 아직 video outpainting paper는 CVPR workshop paper 이외에는 존재하지 않음
      - Video generation research paper의 접근 방식을 따라가도 괜찮겠다 생각

2022 CVPR Video outpainting paper

감사합니다

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - APPENDIX
    - Evaluation metric
      - MSE
        - ✓I번째 학습 데이터의 GT와 predict값
      - PSNR
        - ✓영상 내 신호가 가질 수 있는 최대 신호 잡음의 비율
        - ✓화질이 얼마나 손실되었는지 평가
        - ✓S는 pixel의 최대값
      - SSIM
        - ✓인간의 시각적 화질 차이 평가
        - ✓Luminance, Contrast, Structural 평가
        - ✓A = 원본 이미지
        - ✓B = 왜곡 이미지

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

$$PSNR = 10\log\frac{s^2}{MSE}$$

$$SSIM(A,B) = l(A,B)c(A,B)s(A,B) = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)}$$

2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - APPENDIX
    - Evaluation metric

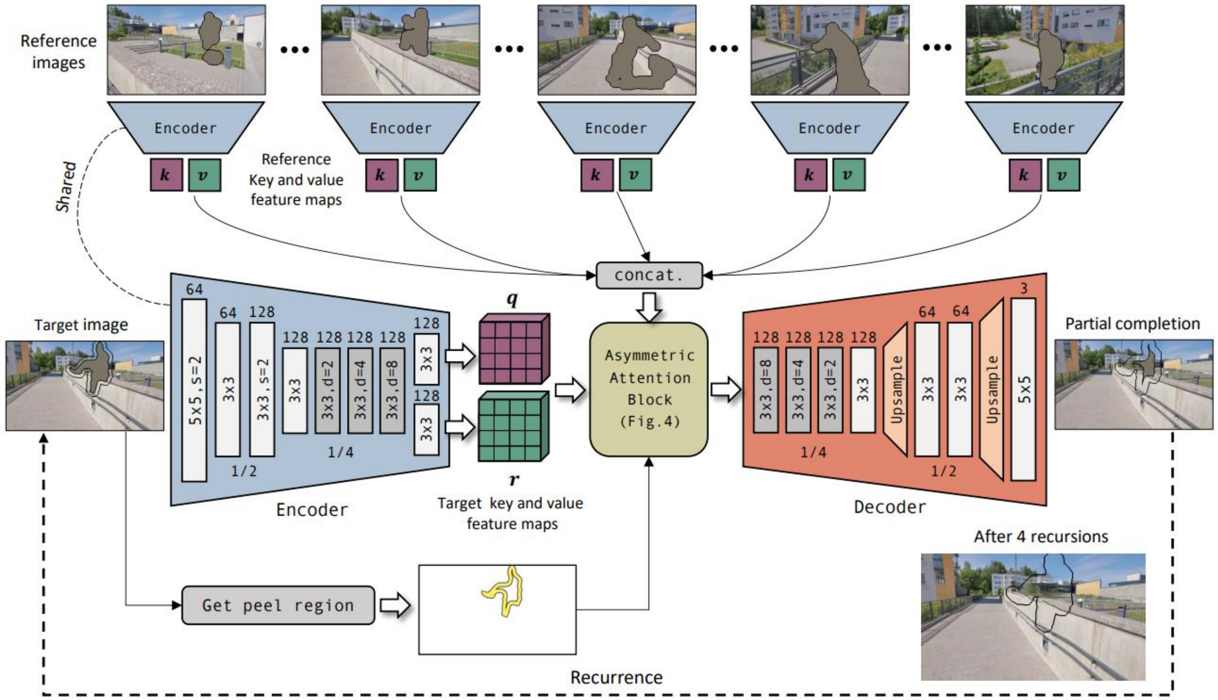
$$LPIPS = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \| w^l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw}) \|_2^2$$

      - LPIPS
        - ✓유사도를 사람의 인식에 기반하여 측정
        - ✓두 이미지 x, x0가 주어졌을 때, layer l에서의 activation map을 얻어 Euclidean distance 계산한 후 w^l로 scaling한 다음 channel wise averaging 한 값을 l에 대해 평균
      - FVD
        - ✓X, Y는 두개의 다변량 정규분포
        - ✓Tr 은 행렬의 대각합(linear algebra)
        - ✓$\sum X \sum Y$: 공분산 행렬(covariance matrix)

$$\mathrm{FID} = ||\mu_X - \mu_Y||^2 - \mathrm{Tr}(\textstyle\sum_X + \sum_Y - 2 \quad \sum_X \sum_Y)$$

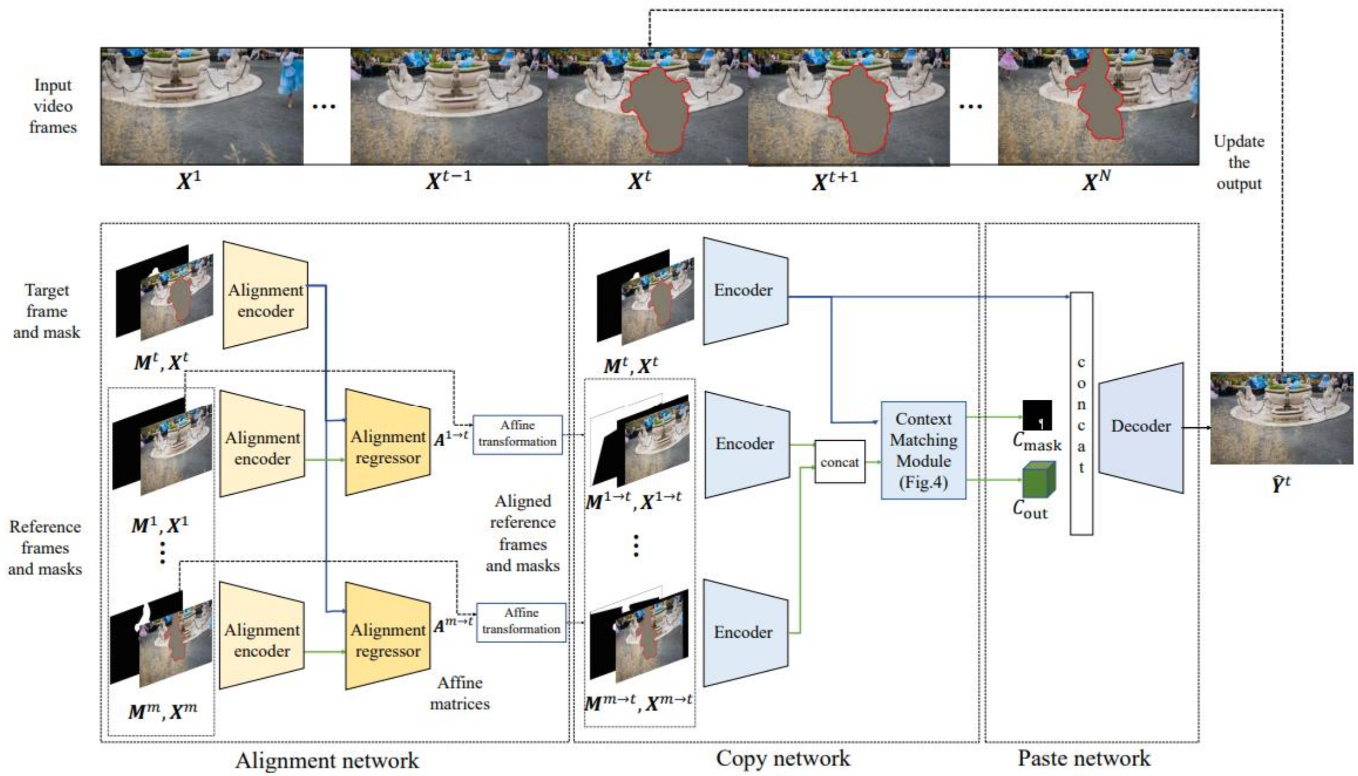
2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - APPENDIX
    - Onion Peel network ICCV_2019



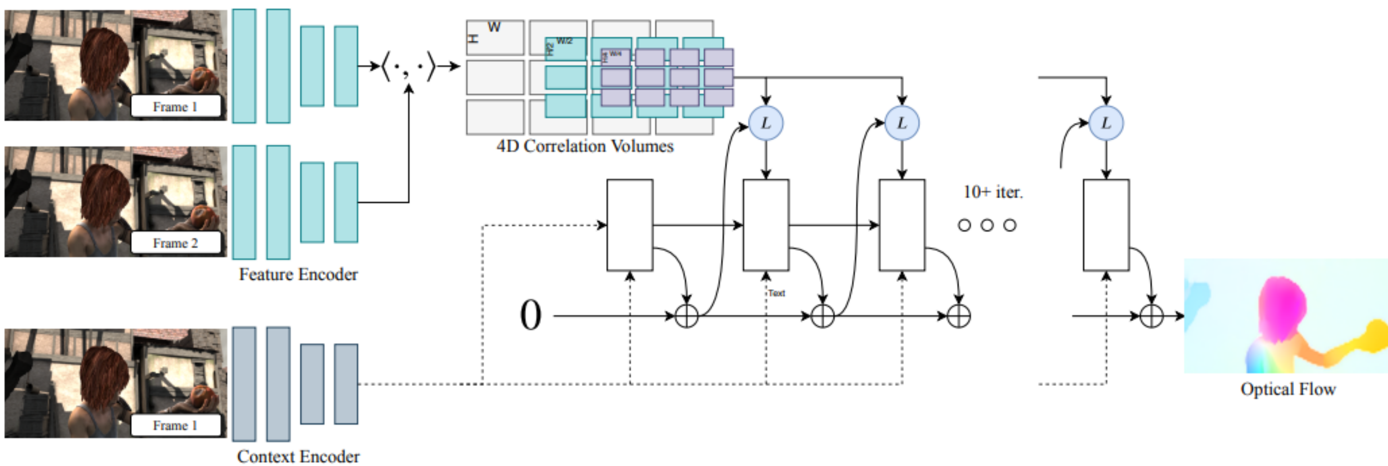
2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - APPENDIX
    - Copy and Paste network ICCV_2019



2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - APPENDIX
    - RAFT ECCV_2020 (Best Paper)
      - 크게 3가지 구조
        - ✓Feature extractor : motion을 위한 feature를 추출
        - ✓Correlation volume : 내적을 통해 C^k : $[H\ W\ \frac{H}{2^k}\ \frac{W}{2^k}]$ 를 구함
        - ✓GRU structure(update operator) : f_k+1= f_k + Δf_k 과정을 반복



2022 CVPR Video outpainting paper

- Complete and Temporally consistent video outpainting
  - APPENDIX
    - RAFT ECCV_2020 (Best Paper)
      - Correlation volume