

RAFT 후속 논문 SEA RAFT Optical flow 추정 기법의 진화

／
SEA_RAFT의 구조와 성능

목 차

01	논문 소개
02	파이프라인
03	실험 결과
04	Ablation
05	결론

RAFT

ECCV 2020년 Princeton University 논문

“RAFT: Recurrent All-Pairs Field Transforms for Optical Flow”

저자

Zachary Teed, Jia Deng (Princeton University)

학회

ECCV 2020 (European Conference on Computer Vision) - Best Paper Award 수상

핵심

모든 픽셀 쌍의 관계를 반복적으로 갱신하여 Optical Flow를 추정하는 방법

SEA RAFT

ECCV 2024년 Princeton University 논문

“SEA-RAFT: Simple, Efficient, Accurate RAFT for Optical Flow”

저자

Yihan Wang, Lahav Lipson, and Jia Deng (Princeton University)

학회

ECCV 2024 (European Conference on Computer Vision)

핵심

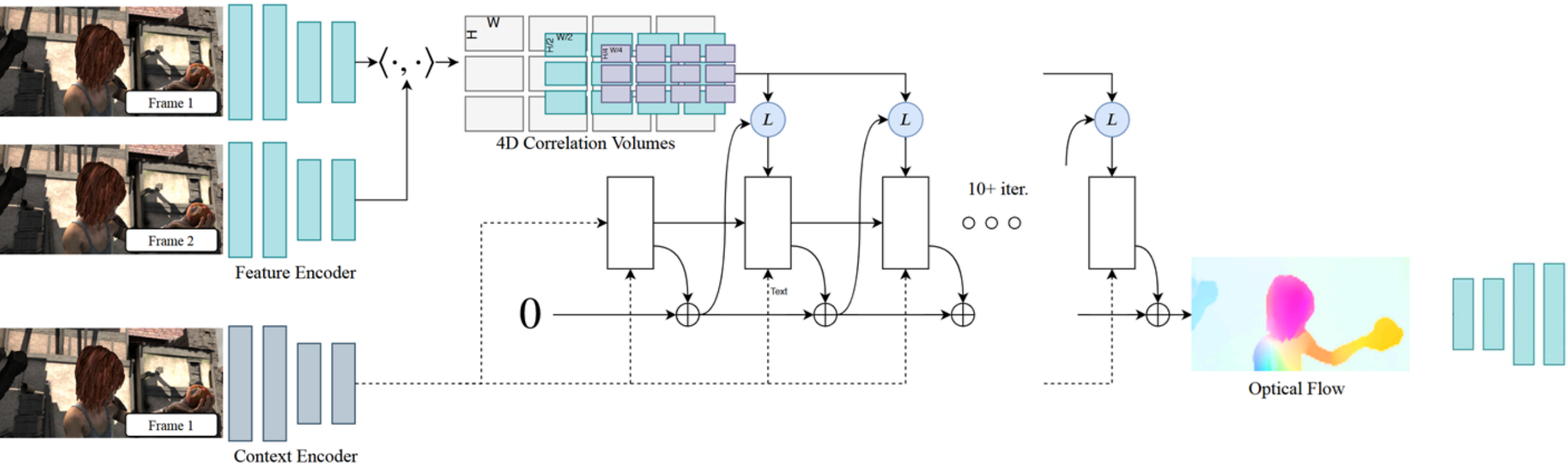
RAFT의 구조를 단순화하면서도 새로운 확률 기반 Loss와 초기 Flow 예측을 도입해
정확도와 효율을 동시에 향상시킨 모델

파이프라인

RAFT

ECCV 2020년 Princeton University 논문

RAFT 파이프라인



학습 데이터셋

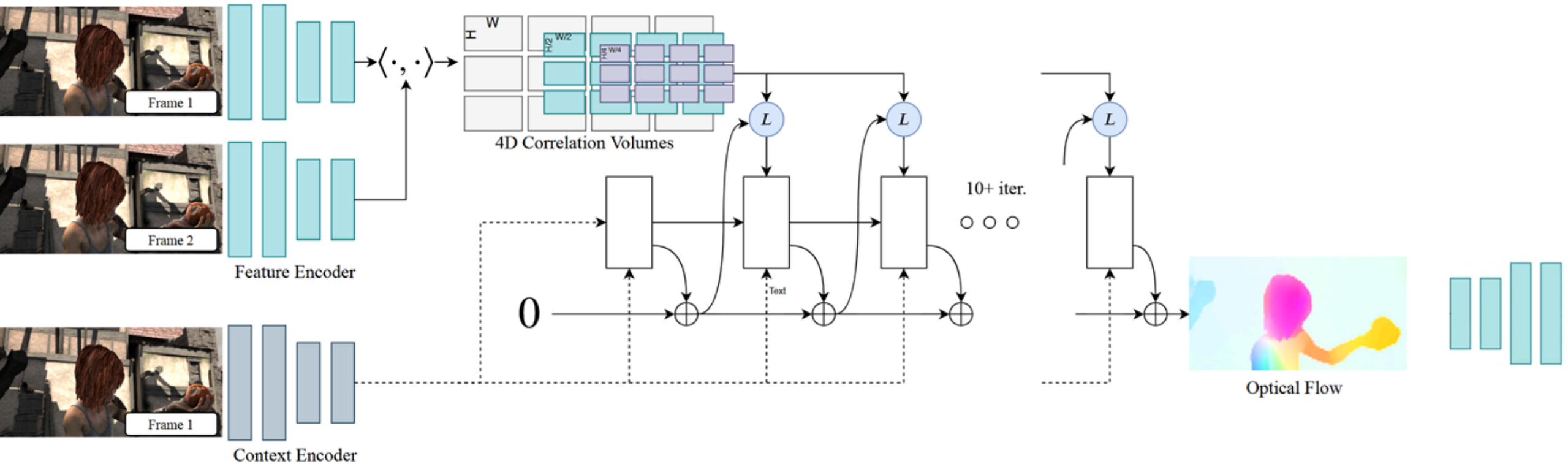


RAFT

ECCV 2020년 Princeton University 논문

RAFT 파이프라인

- 1. 특징 추출
- 2. 4D Lookup table 생성
- 3. Separable GRU
- 4. Upsampling
- 5. 지도학습



학습 데이터셋



Ground Truth Optical Flow

지도학습



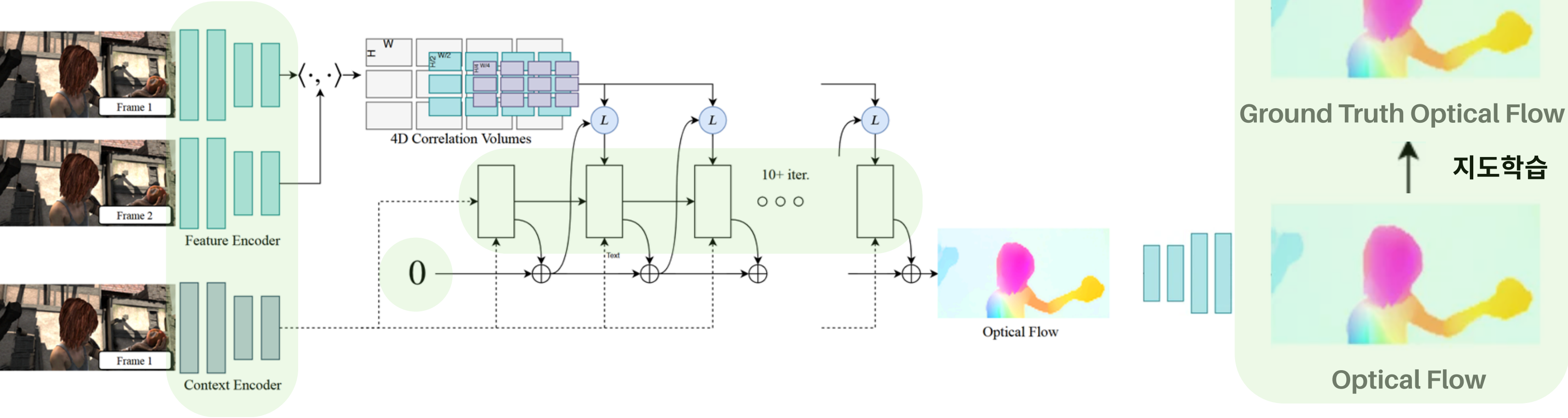
Optical Flow

SEA-RAFT

ECCV 2024년 Princeton University 논문

전체 파이프라인

총 5개의 위치에 큰 변경점이 있다.



SEA-RAFT

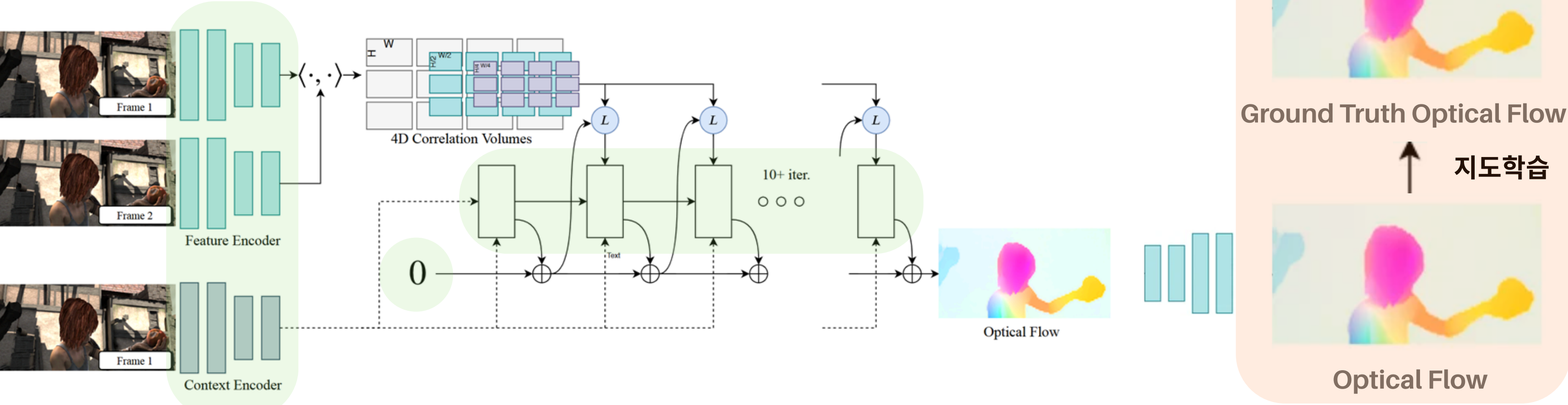
ECCV 2024년 Princeton University 논문

전체 파이프라인

총 5개의 위치에 큰 변경점이 있다.

1. 지도학습 손실함수의 변경

학습 데이터셋



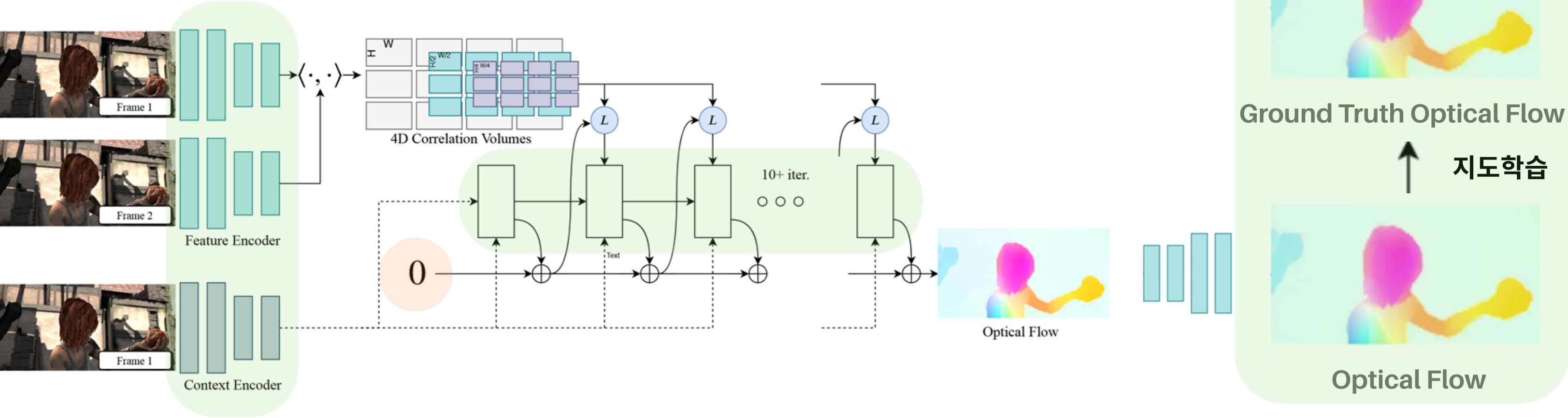
SEA-RAFT

ECCV 2024년 Princeton University 논문

전체 파이프라인

총 5개의 위치에 큰 변경점이 있다.

- 1. 지도학습 손실함수의 변경
- 2. 제일 처음 초기화 값의 변경



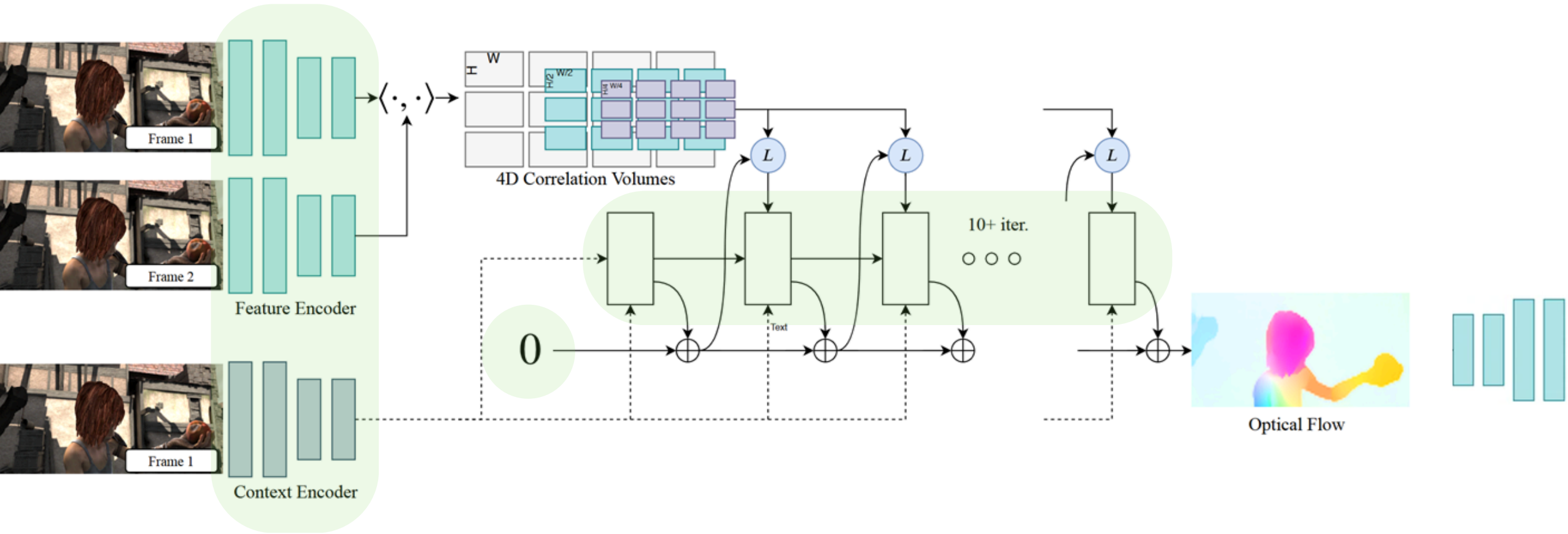
SEA-RAFT

ECCV 2024년 Princeton University 논문

전체 파이프라인

총 5개의 위치에 큰 변경점이 있다.

- 1. 지도학습 손실함수의 변경
- 2. 제일 처음 초기화 값의 변경
- 3. 데이터셋의 추가



학습 데이터셋



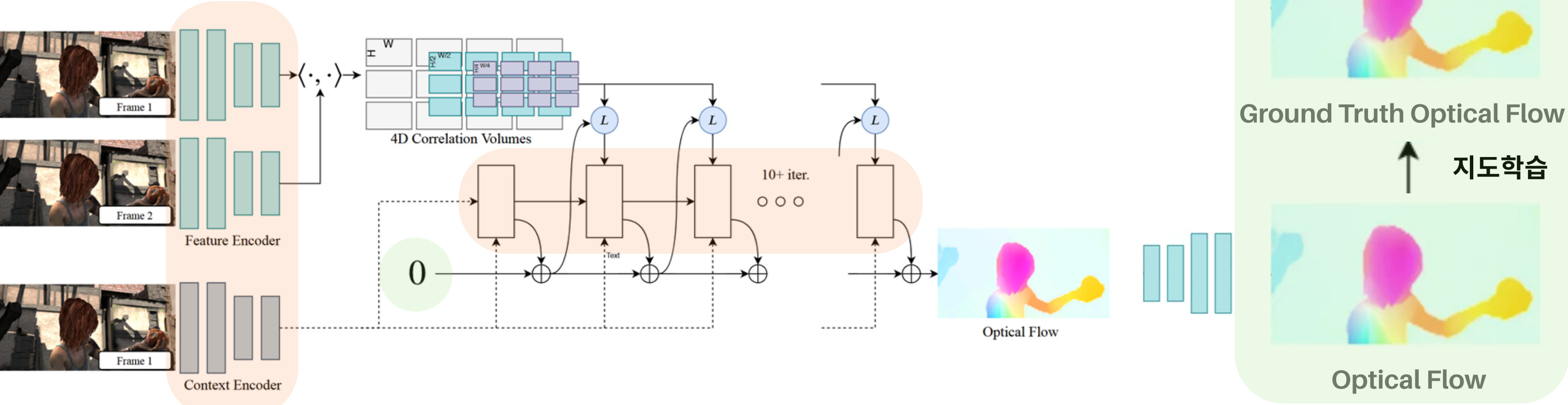
SEA-RAFT

ECCV 2024년 Princeton University 논문

전체 파이프라인

총 5개의 위치에 큰 변경점이 있다.

- 1. 지도학습 손실함수의 변경
- 2. 제일 처음 초기화 값의 변경
- 3. 데이터셋의 추가
- 4. 구조의 간소화



파이프라인

지도학습 손실 함수의 변경

SEA-RAFT

ECCV 2024년 Princeton University 논문

기존 RAFT 손실함수

N: GRU 블록 반복수

$$\mathcal{L} = \sum_{i=1}^N \gamma^{N-i} ||\mathbf{f}_{gt} - \mathbf{f}_i||_1$$

예측한 Flow와 GT Flow의 거리를 최소화하는 것이 목표

픽셀 값의
오차 범위

RAFT는 Ground Truth Flow를 이용한 지도학습을 수행하며,
GT가 존재하는 픽셀을 기준으로 Loss를 계산한다

기존의 Loss는 하나의 정답만을 요구하고 있어서
하나의 Flow를 강제로 맞추도록 설계되어있다

SEA-RAFT

ECCV 2024년 Princeton University 논문

기존 RAFT 손실함수

N: GRU 블록 반복수

$$\mathcal{L} = \sum_{i=1}^N \gamma^{N-i} ||\mathbf{f}_{gt} - \mathbf{f}_i||_1$$

예측한 Flow와 GT Flow의 거리를 최소화하는 것이 목표

픽셀 값의
오차 범위

RAFT는 Ground Truth Flow를 이용한 지도학습을 수행하며,
GT가 존재하는 픽셀을 기준으로 Loss를 계산한다

기존의 Loss는 하나의 정답만을 요구하고 있어서
하나의 Flow를 강제로 맞추도록 설계되어있다

SEA RAFT의 Loss는 이를 해결하기 위해 Flow를 확률 분포로 정의하고,
일반적인 픽셀과 애매한 픽셀을 나눠서 계산합니다.

SEA-RAFT

ECCV 2024년 Princeton University 논문



SEA RAFT 손실함수 유도 과정

확률적 Loss로 접근

$$L_{prob} = -\log p_{\theta}(\mu = \mu_{gt} | I_1, I_2)$$

GT Flow가 발생할 확률을 최대화하는 방향으로 학습
Ground Truth Flow가 나올 확률이 높아질수록 Loss는 작아진다

SEA-RAFT

ECCV 2024년 Princeton University 논문

SEA RAFT 손실함수 유도 과정

확률적 Loss로 접근

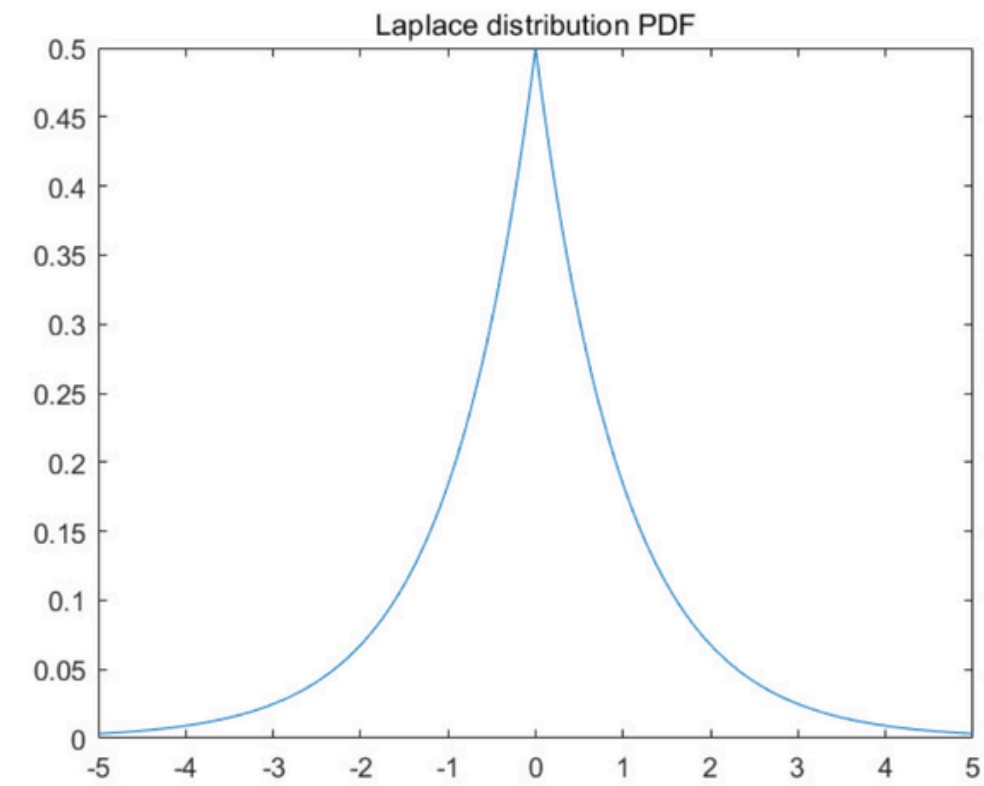
$$L_{prob} = -\log p_{\theta}(\mu = \mu_{gt} | I_1, I_2)$$

GT Flow가 발생할 확률을 최대화하는 방향으로 학습
Ground Truth Flow가 나올 확률이 높아질수록 Loss는 작아진다

Laplace/Gaussian 분포로 확률 접근 => Laplace의 우세

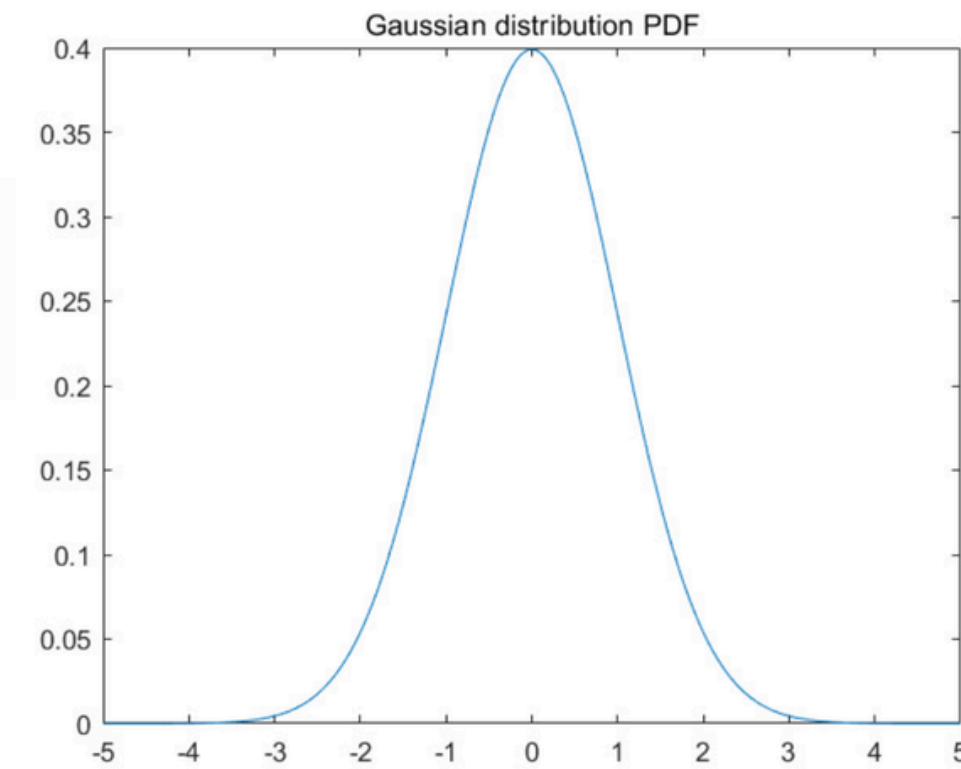
$$p(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

Laplace 분포 수식



$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Gaussian 분포 수식



SEA-RAFT

ECCV 2024년 Princeton University 논문

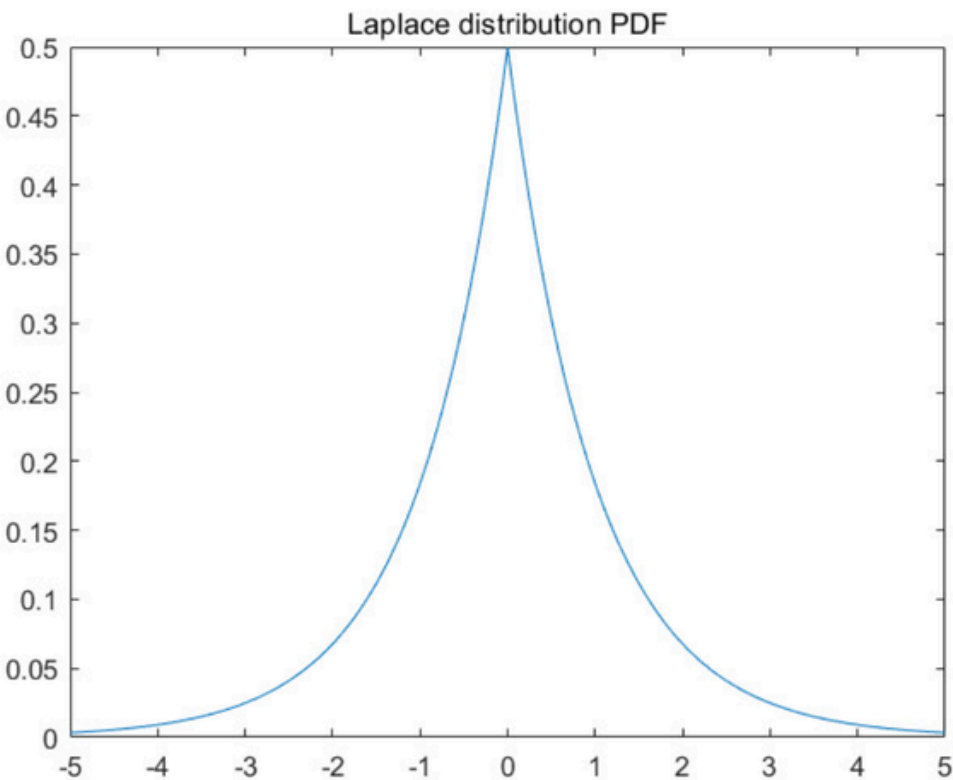
SEA RAFT 손실함수 유도 과정

확률적 Loss로 접근

$$L_{prob} = -\log p_{\theta}(\mu = \mu_{gt} | I_1, I_2)$$

GT Flow가 발생할 확률을 최대화하는 방향으로 학습
Ground Truth Flow가 나올 확률이 높아질수록 Loss는 작아진다

손실 함수 유도



$$p(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

Laplace 분포 수식

$$\rightarrow L_{prob} = -\log\left(\frac{1}{2b} e^{-\frac{|x-\mu|}{b}}\right) = \log(2b) + \frac{|x-\mu|}{b}$$

Lprob에 Laplace 수식을 넣음

SEA-RAFT

ECCV 2024년 Princeton University 논문

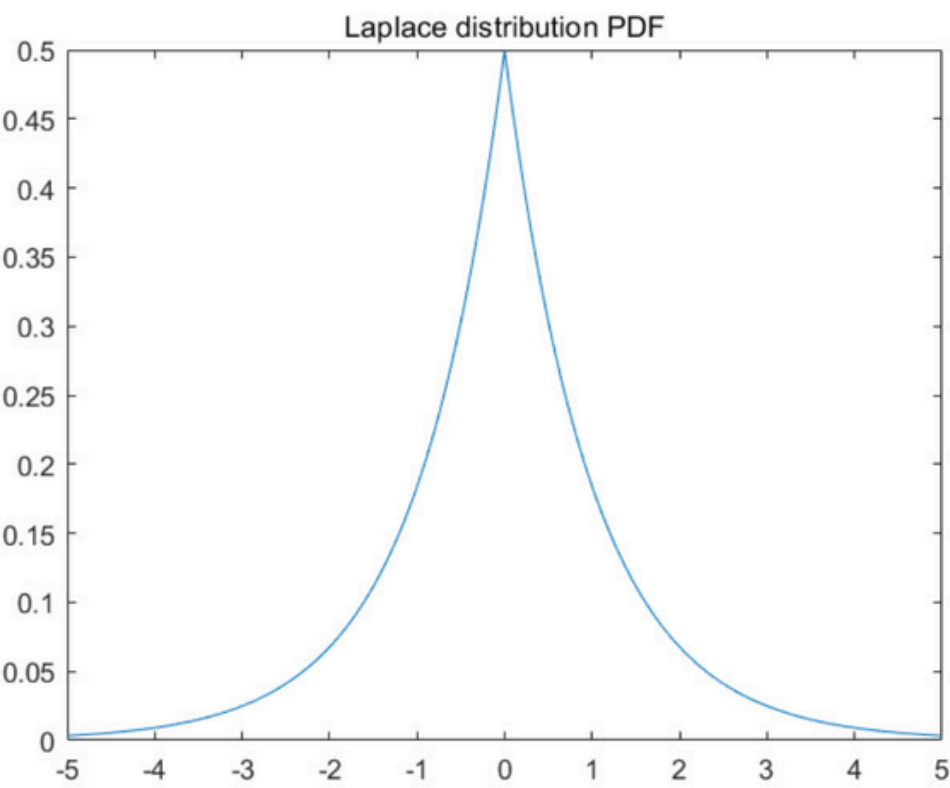
SEA RAFT 손실함수 유도 과정

확률적 Loss로 접근

$$L_{prob} = -\log p_{\theta}(\mu = \mu_{gt} | I_1, I_2)$$

GT Flow가 발생할 확률을 최대화하는 방향으로 학습
Ground Truth Flow가 나올 확률이 높아질수록 Loss는 작아진다

손실 함수 유도



$$p(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

Laplace 분포 수식

$$\rightarrow L_{prob} = -\log\left(\frac{1}{2b} e^{-\frac{|x-\mu|}{b}}\right) = \log(2b) + \frac{|x-\mu|}{b}$$

Lprob에 Laplace 수식을 넣음

$$L_{u,v} = \frac{L_{u,v,x} + L_{u,v,y}}{2} = \frac{1}{2} \left[2\log(2b(u,v)) + \frac{|\mu_{gt,x} - \mu_x| + |\mu_{gt,y} - \mu_y|}{b(u,v)} \right] = \log(2b(u,v)) + \frac{|\mu_{gt,x} - \mu_x| + |\mu_{gt,y} - \mu_y|}{2b(u,v)}$$

한 픽셀에 대한 Loss

SEA-RAFT

ECCV 2024년 Princeton University 논문

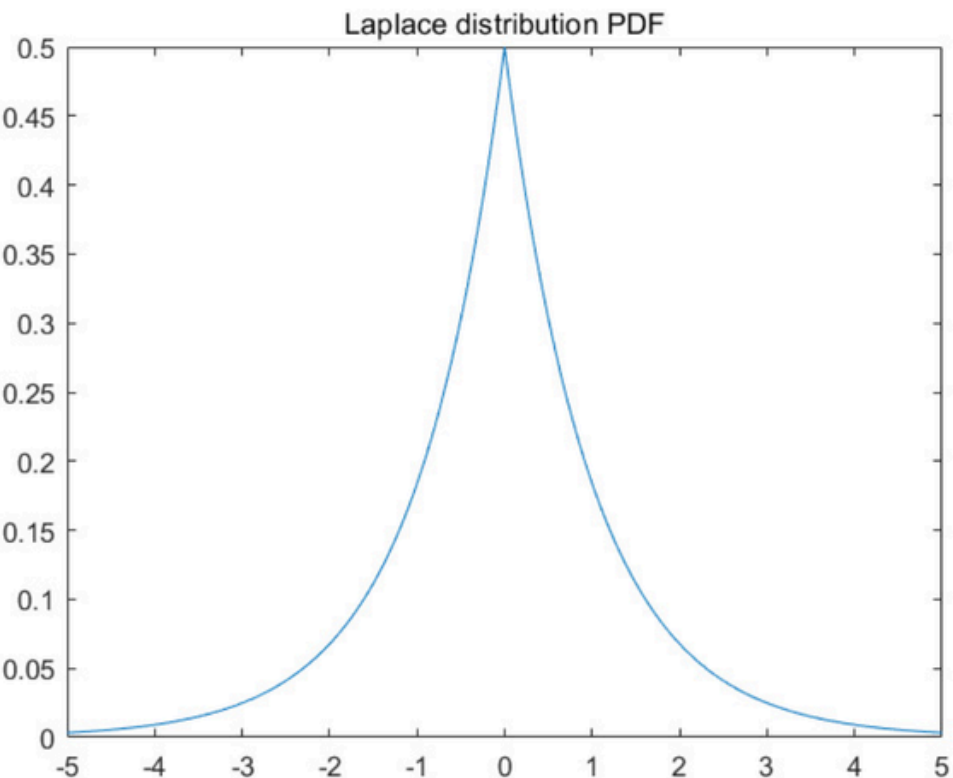
SEA RAFT 손실함수 유도 과정

확률적 Loss로 접근

$$L_{prob} = -\log p_{\theta}(\mu = \mu_{gt} | I_1, I_2)$$

GT Flow가 발생할 확률을 최대화하는 방향으로 학습
Ground Truth Flow가 나올 확률이 높아질수록 Loss는 작아진다

손실 함수 유도



$$p(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

Laplace 분포 수식

$$\rightarrow L_{prob} = -\log\left(\frac{1}{2b} e^{-\frac{|x-\mu|}{b}}\right) = \log(2b) + \frac{|x-\mu|}{b}$$

Lprob에 Laplace 수식을 넣음

$$L_{u,v} = \frac{L_{u,v,x} + L_{u,v,y}}{2} = \frac{1}{2} \left[2\log(2b(u,v)) + \frac{|\mu_{gt,x} - \mu_x| + |\mu_{gt,y} - \mu_y|}{b(u,v)} \right] = \log(2b(u,v)) + \frac{|\mu_{gt,x} - \mu_x| + |\mu_{gt,y} - \mu_y|}{2b(u,v)}$$

한 픽셀에 대한 Loss

$$L_{Lap} = \frac{1}{HW} \sum_u \sum_v \left(\log 2b(u,v) + \frac{\|\mu_{gt}(u,v) - \mu(u,v)\|_1}{2b(u,v)} \right)$$

모든 픽셀에 대해 평균 Loss

SEA-RAFT

ECCV 2024년 Princeton University 논문

SEA RAFT 손실함수 유도 과정

$$L_{Lap} = \frac{1}{HW} \sum_u \sum_v \left(\log 2b(u, v) + \frac{\|\mu_{gt}(u, v) - \mu(u, v)\|_1}{2b(u, v)} \right)$$

L lap의 문제점 1 : 하나의 Laplace 분포로는 힘들수 있다

L lap의 문제점 2 : b에 대하여 log를 취하면 음수를 입력받을 수 없다

$$L_{MoL} = -\frac{1}{2HW} \sum_u \sum_v \sum_{d \in \{x, y\}} \log \left[\alpha \cdot \frac{e^{-\frac{|x-\mu|}{e^{\beta_1}}}}{2e^{\beta_1}} + (1 - \alpha) \cdot \frac{e^{-\frac{|x-\mu|}{e^{\beta_2}}}}{2e^{\beta_2}} \right]$$

L MoL의 해결점 1 : 두 개의 Laplace 분포를 쓴다.

L MoL의 해결점 2 : 지수 함수 기반의 매개변수로 쓴다.

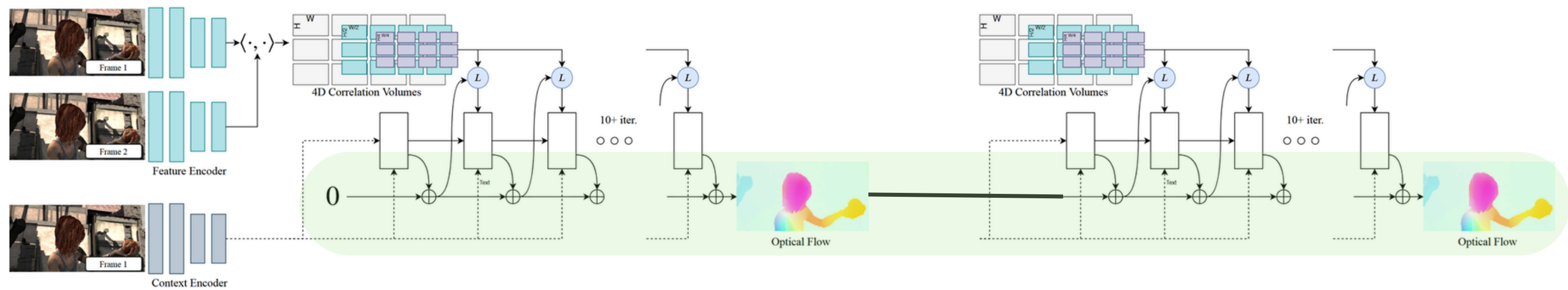
파이프라인

제일 처음 초기화 값의 변경

SEA-RAFT

ECCV 2024년 Princeton University논문

기존 Initial Flow

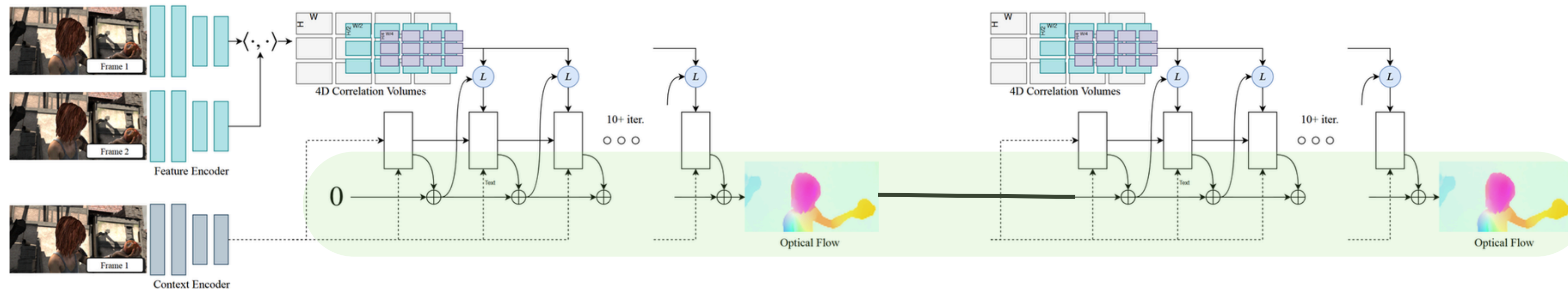


0으로 시작된 값을 점진적으로 업데이트 하는구조

SEA-RAFT

ECCV 2024년 Princeton University 논문

기존 Initial Flow



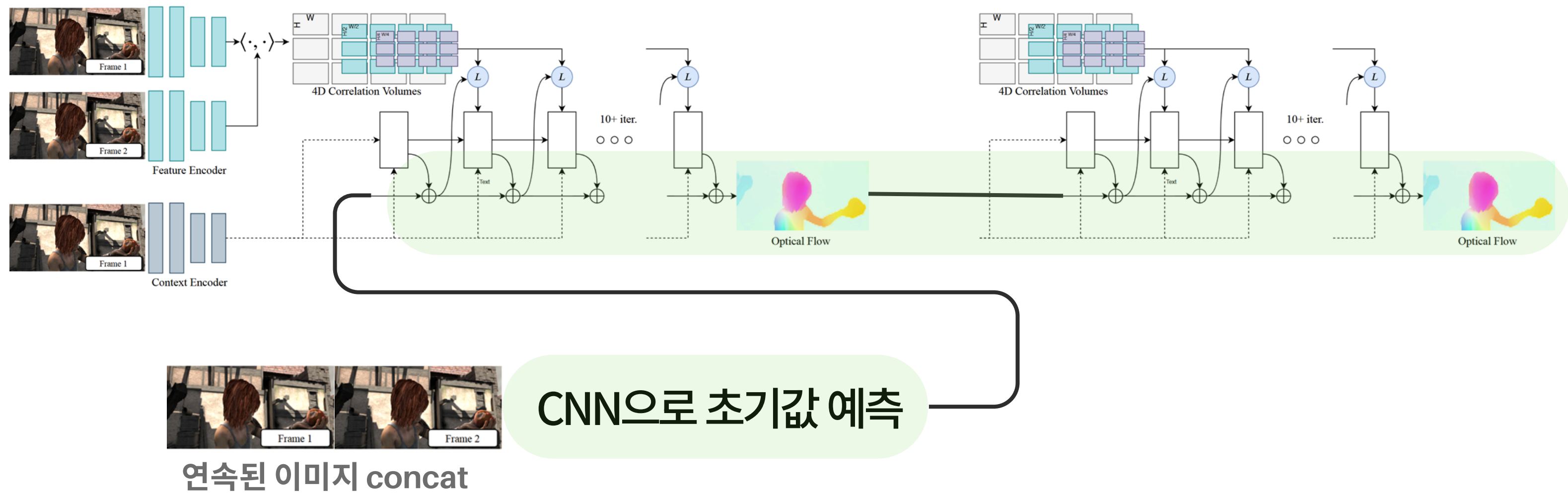
0으로 시작된 값을 점진적으로 업데이트 하는구조

“0조차도 수렴하기 쉽게 바꾸자”

SEA-RAFT

ECCV 2024년 Princeton University논문

Direct Regression of Initial Flow



파이프라인 데이터 셋의 추가

SEA-RAFT

ECCV 2024년 Princeton University 논문

데이터셋 추가

FlyingChairs (C)
FlyingThings3D (T)
Sintel (S)
KITTI (K)
HD1K (H)



TartanAir

- C: 단순한 합성 의자 이미지를 이용해 Optical Flow의 기본 움직임을 학습하는 데이터셋
- T: 다양한 3D 객체와 복잡한 움직임을 포함한 대규모 합성 Optical Flow 데이터셋
- S: 영화 Sintel을 기반으로 제작된 현실적인 조명과 모션 블러를 포함한 벤치마크 데이터셋
- K: 실제 자율주행 환경에서 촬영된 실세계 Optical Flow 성능 평가 데이터셋
- H: 고해상도 실제 영상으로 구성된 복잡한 움직임을 평가하기 위한 데이터셋

다양한 실제와 유사한 환경에서 카메라 움직임(Rigid Flow)
학습을 위한 사전학습 데이터셋

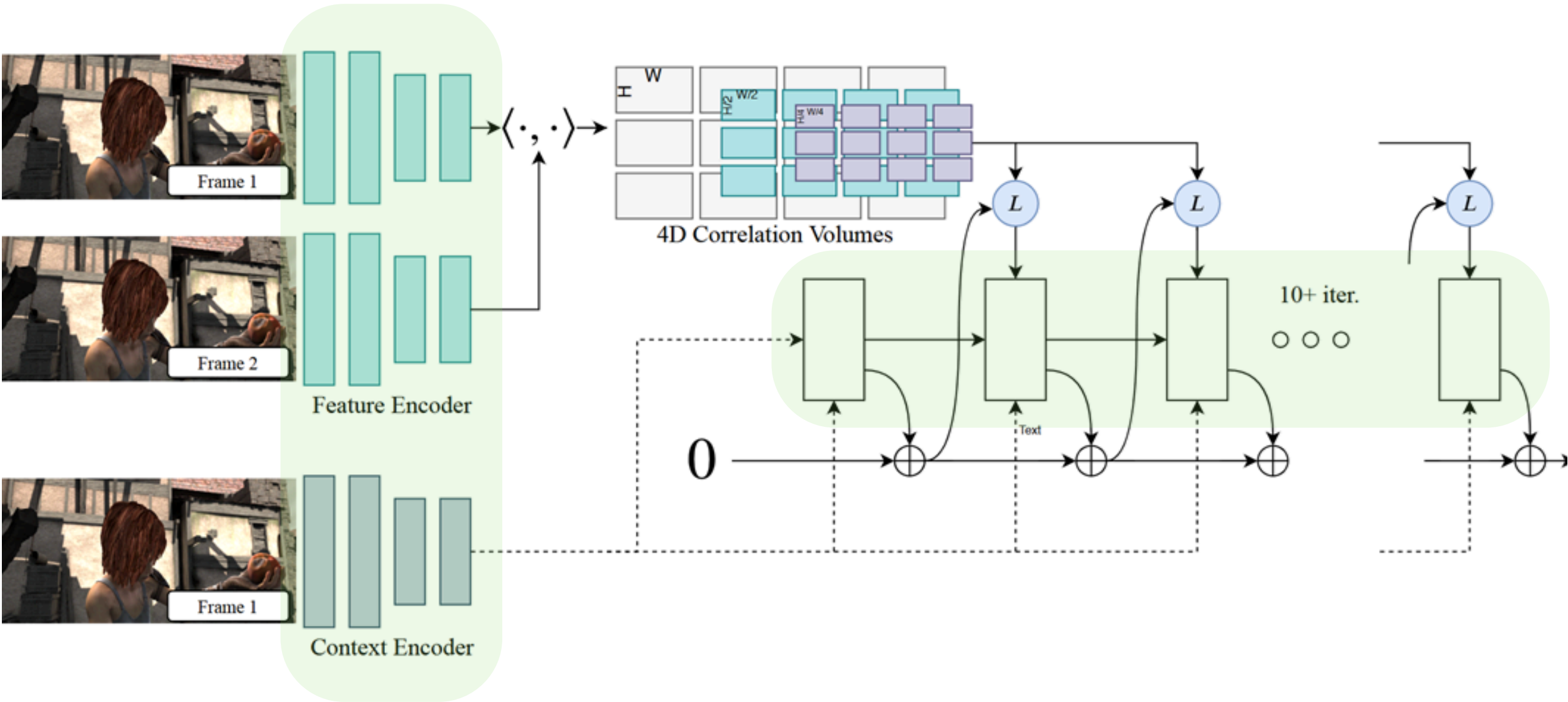
파이프라인 구조의 간소화

SEA-RAFT

ECCV 2024년 Princeton University 논문

구조의 간소화

2가지 변경점이 존재한다.

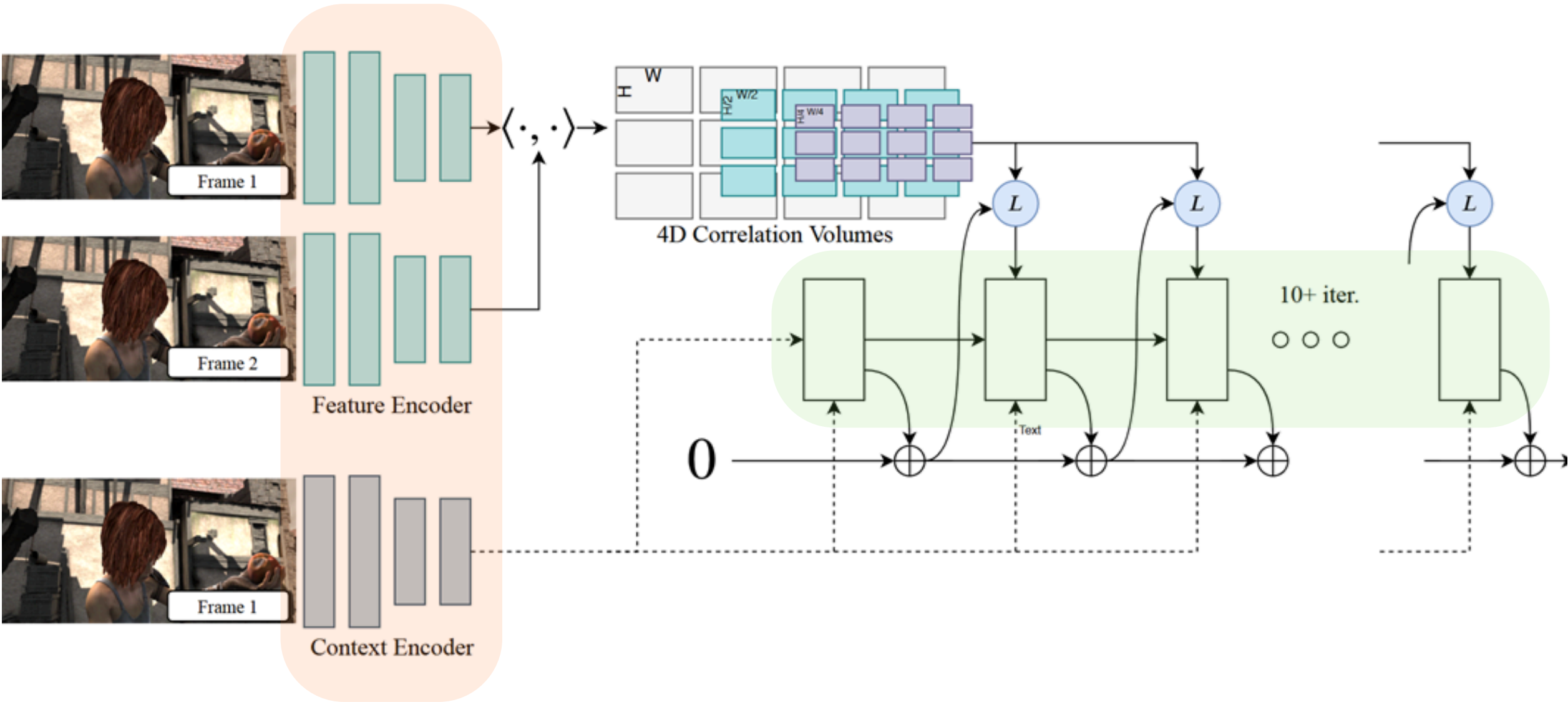


SEA-RAFT

ECCV 2024년 Princeton University논문

구조의 간소화

Feature Encoder와 Context Encoder를 직접 설계 → Truncated ResNet

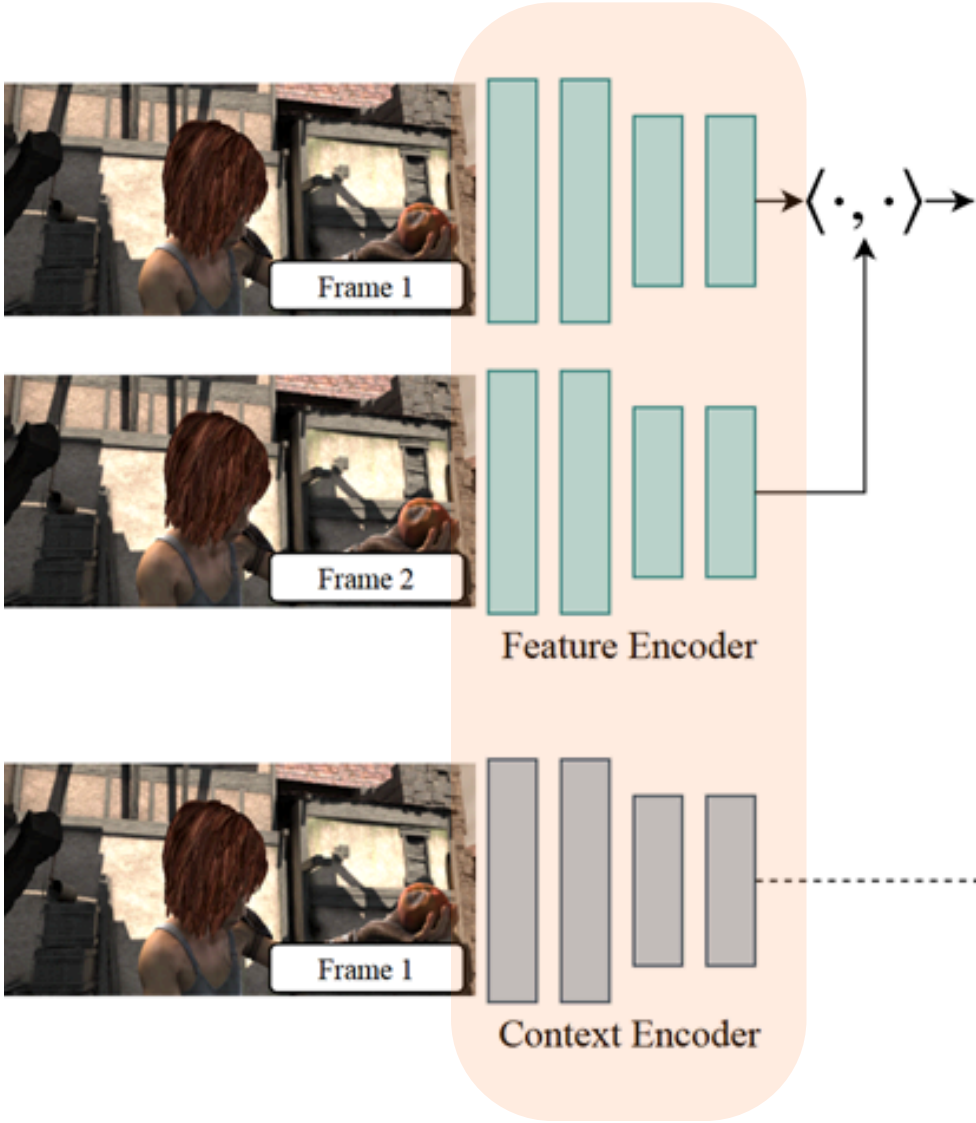


SEA-RAFT

ECCV 2024년 Princeton University 논문

구조의 간소화

Feature Encoder와 Context Encoder를 직접 설계 → Truncated ResNet



장점

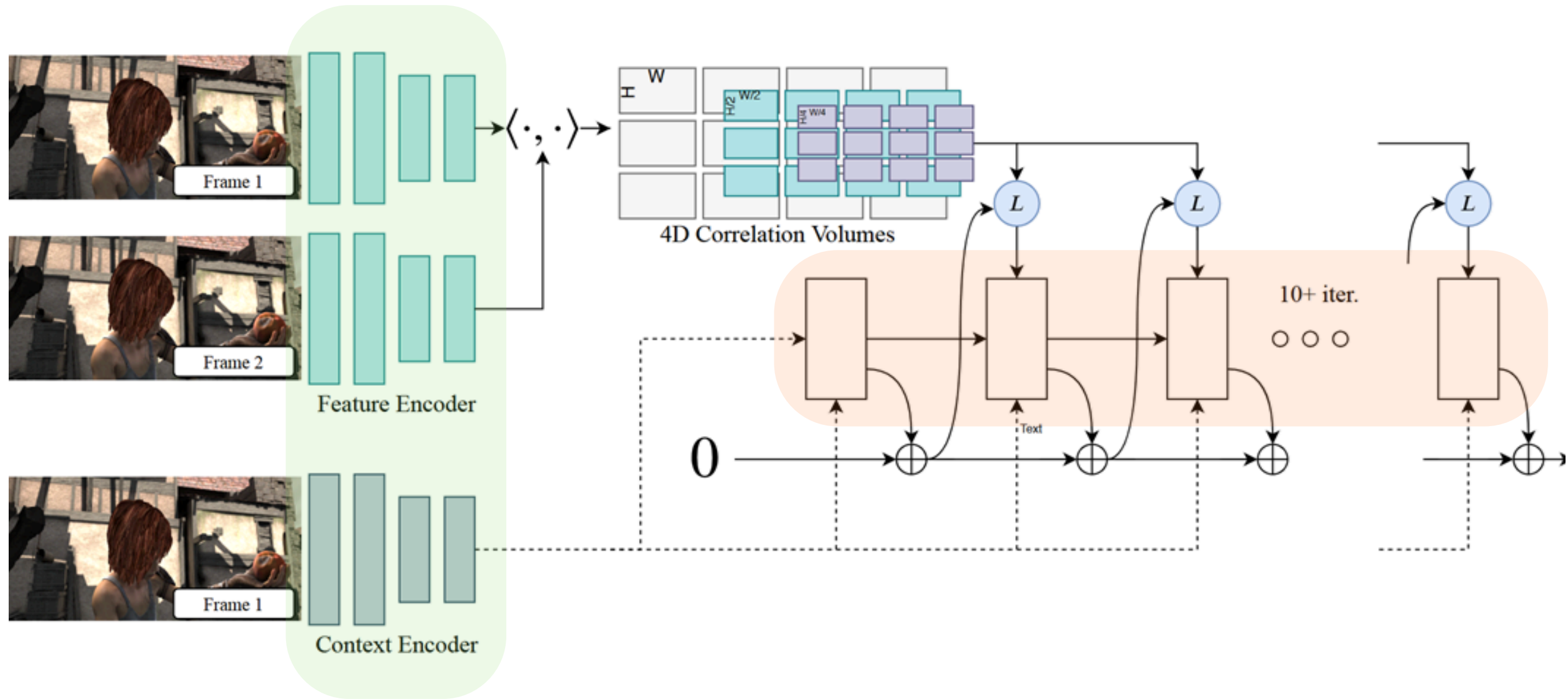
- 1. Custom이 아닌 검증된 ResNet을 사용한다.
- 2. ImageNet 사전학습 데이터를 활용 가능하다.

SEA-RAFT

ECCV 2024년 Princeton University논문

구조의 간소화

ConvGRU → ConvNeXt Block

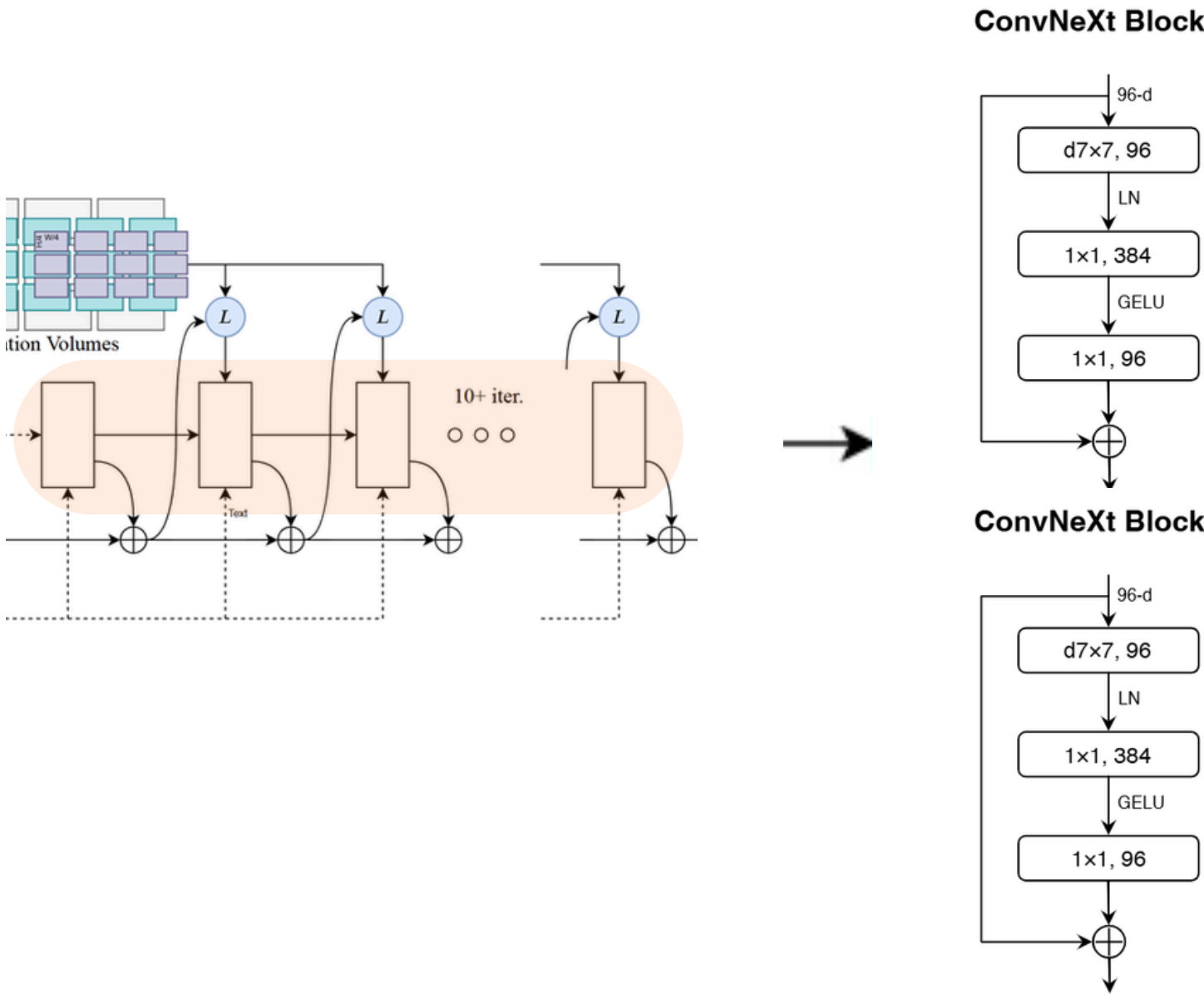


SEA-RAFT

ECCV 2024년 Princeton University 논문

구조의 간소화

ConvGRU → ConvNeXt Block (2개)



장점

- 1. 구조의 단순성 (Update, Reset 등을 안해도 됨)
- 2. 학습 안정성 향상
- 3. 연산량의 감소 → 효율 향상 → Update 속도의 향상

Experiments

결과

2nd

1st

SPRING 벤치마크

현실적인 고해상도 장면에서 Optical Flow 모델의 정확도와 일반화 성능을 평가하기 위한 최신 벤치마크

평가지표

1px:전체 픽셀 중 EPE가 1픽셀을 초과하는 '오답(Outlier)' 픽셀의 비율입니다

EPE:
$$EPE = \sqrt{(u - u_{gt})^2 + (v - v_{gt})^2}$$
각 픽셀의 예측 오차(픽셀 단위)

F1:
$$F1 = \frac{\text{Outlier Pixels}}{\text{Valid Ground Truth Pixels}} \times 100\%$$
Outlier: EPE > 3px 이고 상대오차 > 5%

WAUC: 각 임계값의 픽셀마다 정확한 픽셀 비율을 측정하고, 작은 범위의 영역에 가중치를 더 부여한 지표

실험환경

C+T+S+K+H 로 기본 학습
Extra Data로 추가 학습

SEA-RAFT가 Spring에서 추가 데이터와 fine-tuning을 활용할 경우 최고 성능을 달성하며, 작은 모델도 높은 경쟁력을 가진다는 것을 보여준다.

Extra Data	Method	Spring (train)			Spring (test)			
		Fine-tune	1px↓	EPE↓	1px↓	EPE↓	F1↓	WAUC↑
	PWC-Net [45]	✗	-	-	82.27*	2.288*	4.889*	45.670*
	FlowNet2 [17]	✗	-	-	6.710*	1.040*	2.823*	90.907*
	RAFT [50]	✗	4.788	0.448	6.790*	1.476*	3.198*	90.920*
	GMA [20]	✗	4.763	0.443	7.074*	0.914*	3.079*	90.722*
	RPKNet [37]	✗	4.472	0.416	4.809	0.657	1.756	92.638
	DIP [68]	✗	4.273	0.463	-	-	-	-
	SKFlow [48]	✗	4.521	0.408	-	-	-	-
	GMFlow [56]	✗	29.49	0.930	10.355*	0.945*	2.952*	82.337*
	GMFlow+ [57]	✗	4.292	0.433	-	-	-	-
	Flowformer [14]	✗	4.508	0.470	6.510*	0.723*	2.384*	91.679*
	CRAFT [44]	✗	4.803	0.448	-	-	-	-
	SEA-RAFT(S)	✗	4.077	0.415	-	-	-	-
	SEA-RAFT(M)	✗	4.060	0.406	-	-	-	-
	MegaDepth [26]	✗	4.504	0.407	-	-	-	-
	YouTube-VOS [59]	✗	4.482	0.447	-	-	-	-
	VIPER [41]	✗	3.577	0.397	5.724*	0.643*	2.189*	92.888*
	TartanAir [52]	✗	4.161	0.410	-	-	-	-
	TartanAir [52]	✗	3.888	0.406	-	-	-	-
	CroCo-Pretrain	✓	-	-	4.565	0.498	1.508	93.660
	CroCo-Pretrain	✓	-	-	5.371	0.475	1.621	92.270
	TartanAir [52]	✓	-	-	3.904	0.377	1.389	94.182
	TartanAir [52]	✓	-	-	3.686	0.363	1.347	94.534

결과

1st

#Benchmark

Sintel: CG 기반 데이터셋
KITTI: 실제 도로 주행 환경을 촬영한 데이터셋

평가지표

Clean: Clean Mean EPE
Final: Final Mean EPE
Fl-epe: train set에서 epe Outlier 판별
Fl-all: 테스트 set에서 epe Outlier 판별

$$EPE = \sqrt{(u - u_{gt})^2 + (v - v_{gt})^2}$$

각 픽셀의 예측 오차(픽셀 단위)

$$F1 = \frac{\text{Outlier Pixels}}{\text{Valid Ground Truth Pixels}} \times 100\%$$

Outlier: EPE > 3px 이고 상대오차 > 5%

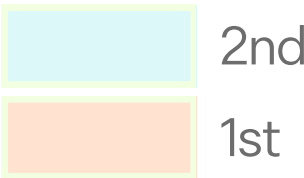
실험환경

C+T로 기본 학습
Extra Data로 추가 학습

Extra Data	Method	Sintel		KITTI	
		Clean↓	Final↓	Fl-epe↓	Fl-all↓
	PWC-Net [45]	2.55	3.93	10.4	33.7
	RAFT [50]	1.43	2.71	5.04	17.4
	GMA [20]	1.30	2.74	4.69	17.1
	SKFlow [48]	1.22	2.46	4.27	15.5
	FlowFormer [14]	1.01	2.40	4.09 [†]	14.7 [†]
	DIP [68]	1.30	2.82	4.29	13.7
	EMD-L [6]	0.88	2.55	4.12	13.5
	CRAFT [44]	1.27	2.79	4.88	17.5
	RPKNet [37]	1.12	2.45	-	13.0
	GMFlowNet [66]	1.14	2.71	4.24	15.4
	SEA-RAFT(M)	1.21	4.04	4.29	14.2
	SEA-RAFT(L)	1.19	4.11	3.62	12.9
Tartan	GMFlow [56]	1.08	2.48	11.2*	28.7*
	GMFlow [56]	-	-	8.70 (-22%)*	24.4 (-15%)*
	SEA-RAFT(S)	1.27	4.32	4.61	15.8
Tartan	SEA-RAFT(S)	1.27	3.74 (-13%)	4.43	15.1
K+H	SEA-RAFT(S)	1.32	2.95 (-32%)	-	-
Tartan+K+H	SEA-RAFT(S)	1.30	2.79 (-35%)	-	-

SEA-RAFT는 학습하지 않은 KITTI에서도 잘 작동하며,
TartanAir와 소량의 실제 데이터가 일반화 성능을 크게 높인다.

결과



#Benchmark

Sintel: CG 기반 데이터셋

KITTI: 실제 도로 주행 환경을 촬영한 데이터셋

평가지표

Clean: Clean Mean EPE

Final: Final Mean EPE

Fl-all: 테스트 set에서 epe Outlier 판별

Fl-bg : Background Flow Outlier

Fl-fg : Foreground Flow Outlier

실험환경

C+T+S+K+H 로 기본 학습

Extra Data로 추가 학습

Extra Data	Method	Sintel		KITTI			Inference Cost	
		Clean↓	Final↓	Fl-all↓	Fl-bg↓	Fl-fg↓	#MACs	Latency
	PWC-Net+ [46]	3.45	4.60	7.72	7.69	7.88	101.3G	23.82ms
	RAFT [50]	1.61*	2.86*	5.10	4.74	6.87	938.2G	140.7ms
	GMA [20]	1.39*	2.47*	5.15	-	-	1352G	183.3ms
	DIP [68]	1.44*	2.83*	4.21	3.86	5.96	3068G	498.9ms
	GMFlowNet [66]	1.39	2.65	4.79	4.39	6.84	1094G	244.3ms
	GMFlow [56]	1.74	2.90	9.32	9.67	7.57	602.6G	138.5ms
	CRAFT [44]	1.45*	2.42*	4.79	4.58	5.85	2274G	483.4ms
	FlowFormer [14]	1.20	2.12	4.68 [†]	4.37 [†]	6.18 [†]	1715G	335.6ms
	SKFlow [48]	1.28*	2.23*	4.85	4.55	6.39	1453G	331.9ms
	GMFlow+ [57]	1.03	2.37	4.49	4.27	5.60	1177G	249.6ms
	EMD-L [6]	1.32	2.51	4.49	4.16	6.15	1755G	OOM
	RPKNet [37]	1.31	2.65	4.64	4.63	4.69	<u>137.0G</u>	183.3ms
VIPER [41]	CCMR+ [18]	<u>1.07</u>	<u>2.10</u>	3.86	3.39	6.21	12653G	OOM
MegaDepth [26]	MatchFlow(G) [8]	1.16*	2.37*	4.63 [†]	4.33 [†]	6.11 [†]	1669G	290.6ms
YouTube-VOS [59]	Flowformer++ [43]	<u>1.07</u>	1.94	4.52 [†]	-	-	1713G	373.4ms
CroCo-Pretrain	CroCoFlow [55]	1.09 [†]	2.44 [†]	<u>3.64</u> [†]	<u>3.18</u> [†]	5.94 [†]	57343G [†]	6422ms [†]
DDVM-Pretrain	DDVM [42]	1.75 [†]	2.48 [†]	3.26 [†]	2.90 [†]	<u>5.05</u> [†]	-	-
TartanAir [52]	SEA-RAFT(M)	1.44	2.86	4.64	4.47	5.49	486.9G	<u>70.96ms</u>
TartanAir [52]	SEA-RAFT(L)	1.31	2.60	4.30	4.08	5.37	655.1G	108.0ms

SEA-RAFT의 가장 큰 장점은 단순히 정확한 것이 아니라, 비슷한 정확도의 모델보다 훨씬 빠르다는 점이다.

Ablations

Ablation

성능하락

1st

	Experiment	Init.	Pre-Training		RNN		Loss Design		#MACs	EPE
			Img [7]	Tar [52]	GRU	#blocks	Type	Params		
Tartan Air 없이	SEA-RAFT (w/o Tar.)	✓	✓	-	-	2	Mixture-of-Laplace	$\beta_1 = 0, \beta_2 \in [0, 10]$	284.7G	0.187
Tartan Air 있어	SEA-RAFT (w/ Tar.)	✓	✓	✓	-	2	Mixture-of-Laplace	$\beta_1 = 0, \beta_2 \in [0, 10]$	284.7G	0.179
ImageNet 있어	w/o Img.	✓	-	-	-	2	Mixture-of-Laplace	$\beta_1 = 0, \beta_2 \in [0, 10]$	284.7G	0.194
Direct Regression of Initial Flow 없이	w/o Direct Reg.	✗	✓	-	-	2	Mixture-of-Laplace	$\beta_1 = 0, \beta_2 \in [0, 10]$	277.3G	0.201
Separable Conv GRU	RAFT GRU	✓	✓	-	✓	-	Mixture-of-Laplace	$\beta_1 = 0, \beta_2 \in [0, 10]$	297.9G	0.189
더 많은 ConvNext Block	More ConvNeXt Blocks	✓	✓	-	-	4	Mixture-of-Laplace	$\beta_1 = 0, \beta_2 \in [0, 10]$	314.7G	0.189
지수함수 미적용 Laplace	Naive Laplace	✓	✓	-	-	2	Naive Single Laplace	$\beta \in [-10, 10]$	284.7G	0.217
							Naive Mixture-of-Laplace	$\beta_1, \beta_2 \in [-10, 10]$		0.248
L1 Loss만 사용	L1	✓	✓	-	-	2	L_1	✗	284.7G	0.206
가우시안 분포 사용	Gaussian	✓	✓	-	-	2	Mixture-of-Gaussian	$\sigma_1 = 1, \sigma_2 = e^{\beta_2}, \beta_2 \in [0, 10]$	284.7G	0.210

Conclusion

Conclusion

Point 1

새로운 Loss

Mixture-of-Laplace Loss
를 적용으로 일반화 성능을
향상시켰습니다.

Point 2

Direct Regression Initial Flow

초기 Optical Flow를 직접
예측하여 반복 횟수를 줄이고
빠른 수렴을 달성했습니다.

Point 3

데이터셋 증강

TartanAir 데이터셋을 통해
다양한 환경에서의 일반화
성능을 향상시켰습니다.

Point 4

구조의 간소화

ResNet과 ConvNeXt로
구조를 단순화하고 정확도,
연산 효율을 향상시켰습니다.

감사합니다.