

# Video Super Resolution

2022년도 하계세미나

---



*Sogang University*

*Vision & Display Systems Lab, Dept. of Electronic Engineering*



*Presented By*

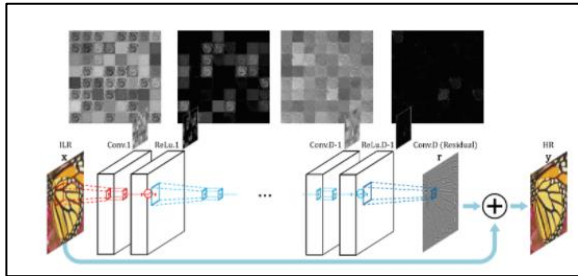
*Yoon Chan Nam*

# Video Super Resolution

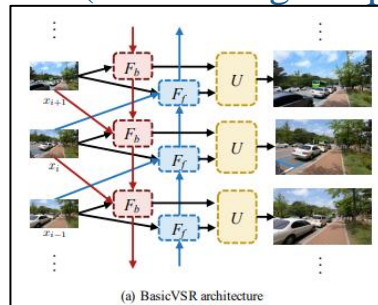
- Super Resolution
  - SISR
  - MISR
  
- VideoSuper Resolution
  - BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond (CVPR 2021)
  - BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment(CVPR 2022)

# Video Super Resolution

- 저해상도(LR) 이미지를 고해상도(HR) 이미지로 변환시키는 task
  - SISR(Single Image Super-Resolution)



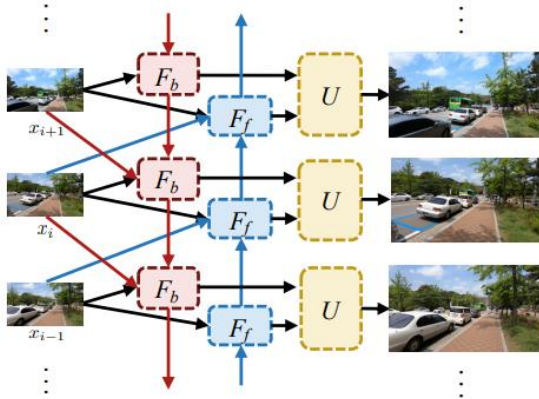
- MISR(Multi-Image Super-Resolution)



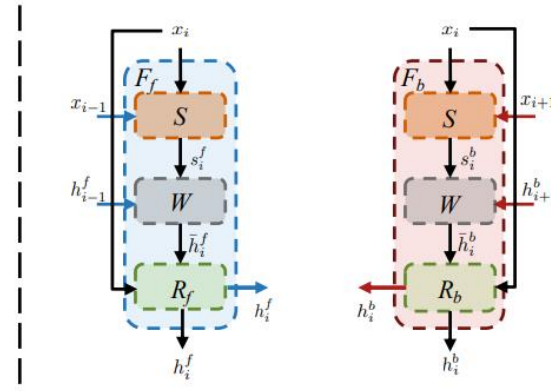
## - Video Super Resolution

- ※ 단일 이미지 SR과 달리 저화질의 Sequential 한 frame을 고화질로 복원
- ※ 주변 frame의 정보를 이용하여 SISR보다 높은 복원 성능

# BasicVSR (CVPR 2021)



(a) BasicVSR architecture



(b) Forward and backward propagation branches

## • Abstract

- VSR이 Sequential 정보를 이용하기 때문에 네트워크 구조가 복잡해지는 문제
- 복잡한 기존 방식을 최소한으로 재설계 하여 복잡성, 성능 개선
  - VSR의 propagation, alignment, fusion, up-sampling의 구성 요소에서 propagation과 alignment요소를 개선

# Basic VSR: Propagation

- Local Propagation

- Sliding window 방식으로, 복원을 위해 local 정보만을 사용
- 접근 가능한 정보가 Local에서 제한됨

- Unidirectional Propagation

- 첫 번째 frame에서 마지막 frame까지 순차적으로 Propagation
- 첫 번째 frame은 자신만 사용, 마지막 frame은 모든 frame의 정보 사용

- Bidirectional Propagation

- 양방향으로 Propagation을 진행
- 위의 두 방식에서의 문제점을 모두 해결

$$\begin{aligned} h_i^b &= F_b(x_i, x_{i+1}, h_{i+1}^b), \\ h_i^f &= F_f(x_i, x_{i-1}, h_{i-1}^f), \end{aligned} \quad (1)$$

# Basic VSR: Propagation

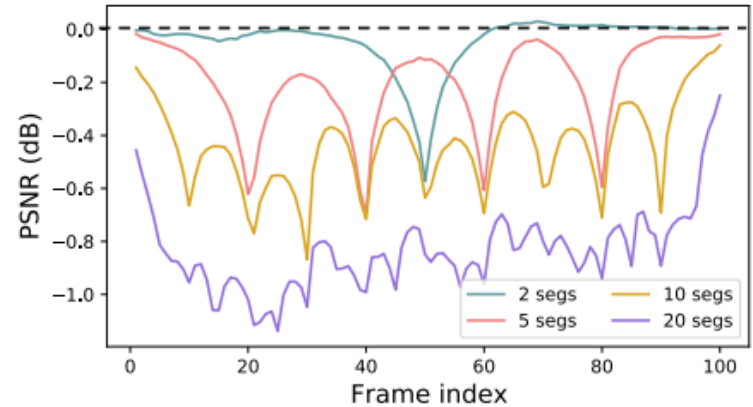
- Local vs Global propagation

- Frame들을 N개의 segment로 나눠 실험

- N Segment 별로 Global propagation

- ※ Local propagation은 segment가 없음

- N = 1 일경우의 PSNR을 기준으로 비교



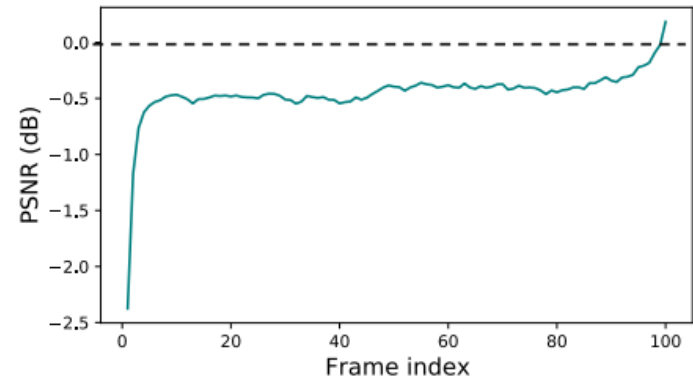
- Unidirectional vs Bidirectional propagation

- Bidirection propagation의 결과를 기준으로 비교

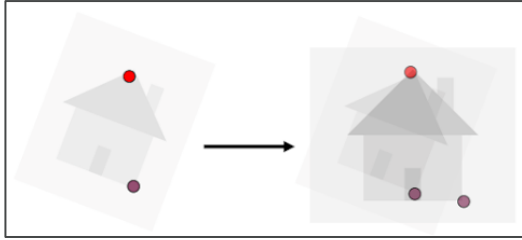
- Unidirection에서는 하위 index의 frame은 정보 부족으로 성능이 낮음

- 상위 index에서는 모든 frame의 정보를 받음

- ※ 평균적으로 Bidirection이 좋음



# Basic VSR: Alignment



- Without Alignment

- Alignment가 되지 않은 image/feature는 information 집계가 어렵고 성능이 떨어짐

- Image Alignment

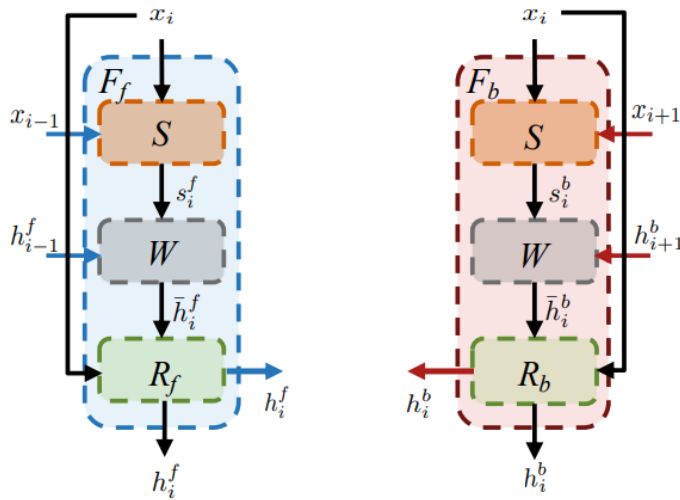
- 복원을 하기 전에 optical flow를 계산하고, 이미지를 warping함
- Feature level에서 alignment를 수행하는 것 보다 성능이 좋지 않음

- Feature Alignment

- 두 프레임 간의 optical flow를 계산하여 이를 offset으로 feature level에서 warping을 수행하여 정렬
- Image alignment 보다 alignment error가 적음
- Without Alignment보다 1.19dB, Image Alignment보다 0.17dB의 성능 향상

# Basic VSR

- Bidirectional Propagation & Feature Alignment



(b) Forward and backward propagation branches

$$h_i^b = F_b(x_i, x_{i+1}, h_{i+1}^b), \quad (1)$$

$$h_i^f = F_f(x_i, x_{i-1}, h_{i-1}^f),$$

$$s_i^{\{b,f\}} = S(x_i, x_{i\pm 1}),$$

$$\bar{h}_i^{\{b,f\}} = W(h_{i\pm 1}^{\{b,f\}}, s_i^{\{b,f\}}), \quad (2)$$

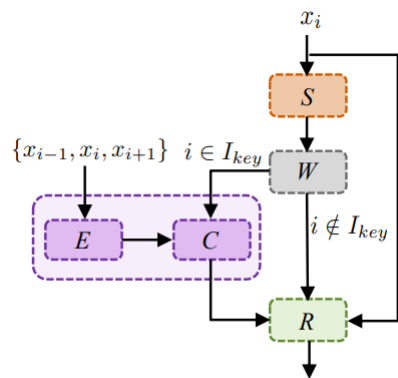
$$h_i^{\{b,f\}} = R_{\{b,f\}}(x_i, \bar{h}_i^{\{b,f\}}),$$

- $S$ :  $x_i$ 와  $x_{i\pm 1}$  간의 optical flow estimation
- $W$ : 이전 frame feature를 현재 frame에 맞게 alignment
- $R$ : aligned feature와 현재 frame을 concatenation하여 residual block 통과



# IconVSR (Information-refill and Coupled propagation)

- Information Refill

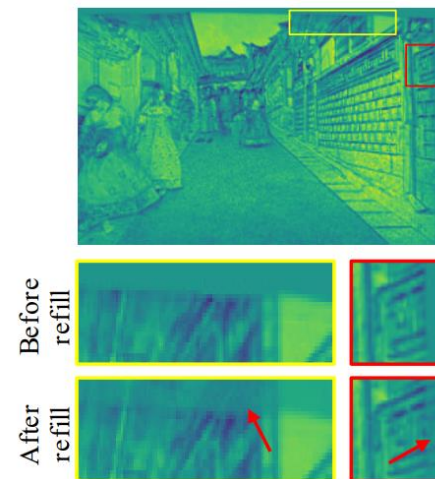


(a) Information-Refill

$$e_i = E(x_{i-1}, x_i, x_{i+1}),$$

$$\hat{h}_i^{\{b,f\}} = \begin{cases} C(e_i, \bar{h}_i^{\{b,f\}}) & \text{if } i \in I_{key}, \\ \bar{h}_i^{\{b,f\}} & \text{otherwise,} \end{cases} \quad (4)$$

$$h_i^{\{b,f\}} = R_{\{b,f\}}(x_i, \hat{h}_i^{\{b,f\}}). \quad (5)$$

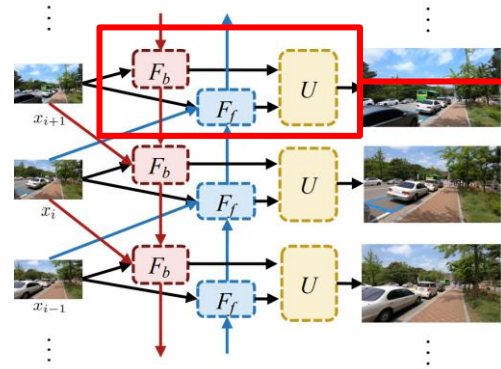


(a) Information-Refill

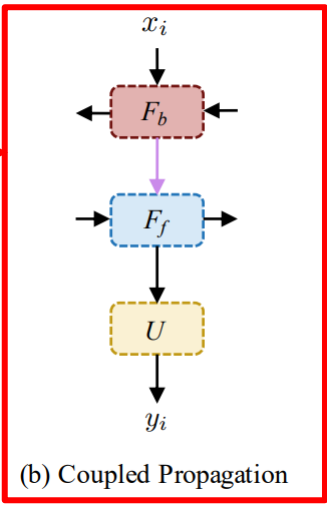
- Long term propagation을 채택한 모델에서 정확하지 않은 정렬은 error를 축적하는 문제
- $\{x_{i-1}, x_i, x_{i+1}\}$ 를 feature extractor E를 통과 시켜 aligned feature와 융합하여 정보를 refill함
- Refill 된 정보로 refine된 feature를 사용하여 정밀도를 올림

# IconVSR (Information-refill and Coupled propagation)

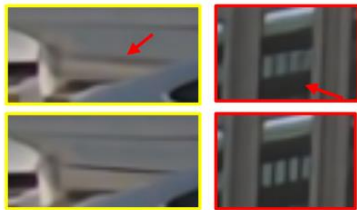
- Coupled Propagation



(a) BasicVSR architecture



(b) Coupled Propagation



Without coupled

With coupled

(b) Coupled Propagation

$$\begin{aligned}
 h_i^b &= F_b(x_i, x_{i+1}, h_{i+1}^b), \\
 h_i^f &= F_f(x_i, x_{i-1}, h_i^b, h_{i-1}^f), \\
 y_i &= U(h_i^f).
 \end{aligned}
 \tag{6}$$

- Backward 와 forward가 상호 연결되는 coupled propagation을 제안
- Backward propagation에서 나온 feature를 forward pass에 추가
- Forward pass에서 이전과 미래 프레임의 정보를 모두 수신

# BasicVSR: Results



	Params (M)	Runtime (ms)	BI degradation			BD degradation		
			REDS4 [23]	Vimeo-90K-T [33]	Vid4 [21]	UDM10 [34]	Vimeo-90K-T [33]	Vid4 [21]
Bicubic	-	-	26.14/0.7292	31.32/0.8684	23.78/0.6347	28.47/0.8253	31.30/0.8687	21.80/0.5246
VESPCN [1]	-	-	-	-	25.35/0.7557	-	-	-
DUF [16]	-	-	-	-	25.88/0.7752	-	-	-
SPMC [29]	-	-	-	-	-	-	-	-
TOFlow [33]	-	-	27.98/0.7990	33.08/0.9054	25.89/0.7651	36.26/0.9438	34.62/0.9212	-
FRVSR [25]	5.1	137	-	-	-	37.09/0.9522	35.64/0.9319	26.69/0.8103
EDVR-M [32]	3.3	118	30.53/0.8699	37.09/0.9446	27.10/0.8186	39.40/0.9663	37.33/0.9484	27.38/0.8329
EDVR [32]	20.6	378	31.09/0.8800	<b>37.61/0.9489</b>	<b>27.35/0.8264</b>	39.89/0.9686	<b>37.81/0.9524</b>	27.45/0.8406
PFNL [34]	3.0	295	29.63/0.8502	36.14/0.9363	26.73/0.8029	38.74/0.9627	-	27.85/0.8503
MuCAN [20]	-	-	30.88/0.8750	37.32/0.9465	-	-	-	27.16/0.8355
TGA [13]	5.8	-	-	-	-	-	37.59/0.9516	27.63/0.8423
RLSP [8]	4.2	49	-	-	-	38.48/0.9606	36.49/0.9403	27.48/0.8388
RSDN [12]	6.2	94	-	-	-	39.35/0.9653	37.23/0.9471	27.92/0.8505
RBPN [9]	12.2	1507	30.09/0.8590	37.07/0.9435	27.12/0.8180	38.66/0.9596	37.20/0.9458	-
BasicVSR (ours)	6.3	63	<b>31.42/0.8909</b>	37.18/0.9450	27.24/0.8251	<b>39.96/0.9694</b>	37.53/0.9498	<b>27.96/0.8553</b>
IconVSR (ours)	8.7	70	<b>31.67/0.8948</b>	<b>37.47/0.9476</b>	<b>27.39/0.8279</b>	<b>40.03/0.9694</b>	<b>37.84/0.9524</b>	<b>28.04/0.8570</b>

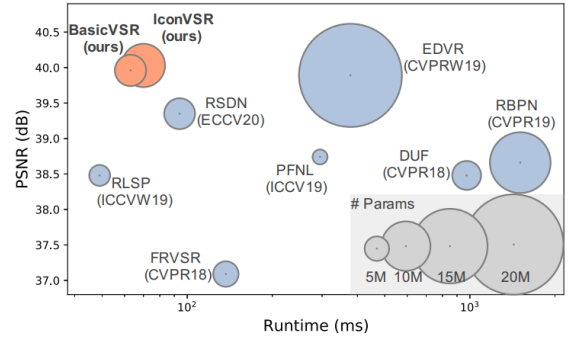
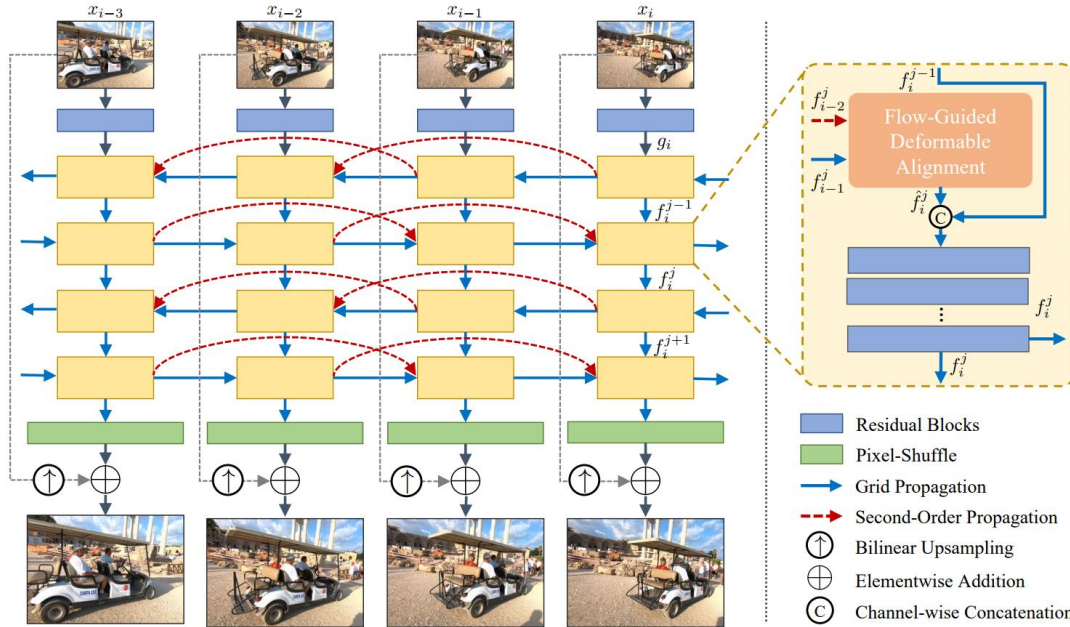


Figure 1. **Speed and performance comparison.** Without bells and whistles, BasicVSR outperforms state-of-the-art methods with high efficiency. Built upon BasicVSR, IconVSR further improves the performance. Comparisons are performed on UDM10 dataset [34].

# BasicVSR++ (CVPR2022)



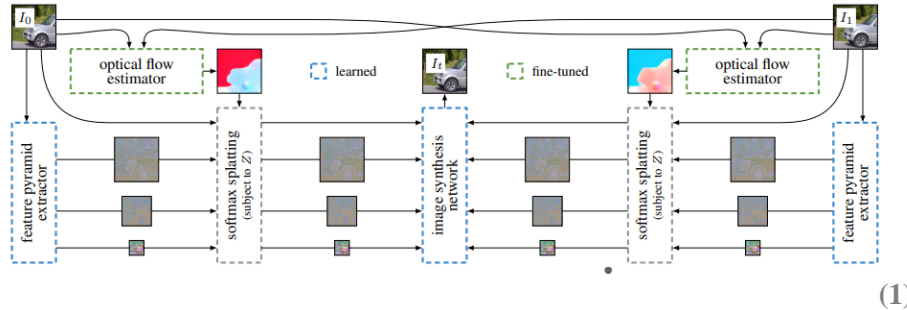
- Abstract

- Second-order Grid Propagation과 Flow-Guided Deformable Alignment를 이용하여 BasicVSR의 Propagation과 Alignment를 강화함

# Basic VSR++: Second-Order Grid Propagation

## • Grid Connection

- Object detection, segmentation, frame interpolation 등과 같은 모델에서 사용
- 일반적으로 Multi-scale로 나뉘어 진행됨
- 양방향으로 시간에 따른 정보를 취합하기 위해 채택

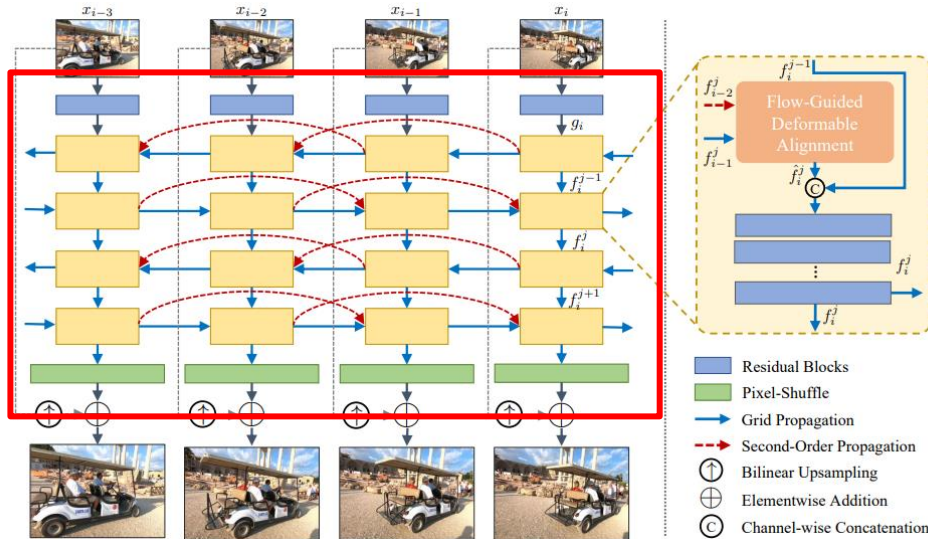


## • Higher-Order Propagation

- Temporal gradient flow 개선을 위해 language model에서 사용
- 시간적 Alignment가 중요한 VSR에서 사용되지 않음

# BasicVSR++: Second-Order Grid Propagation

- Second-Order Grid Propagation



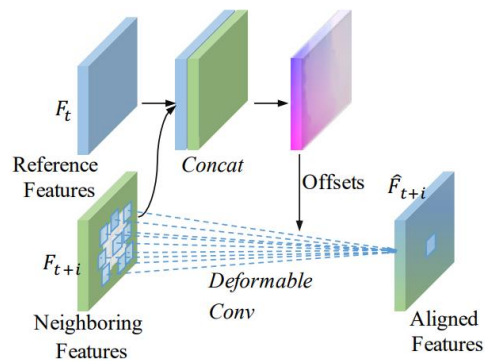
$$\hat{f}_i^j = \mathcal{A} \left( g_i, f_{i-1}^j, f_{i-2}^j, s_{i \rightarrow i-1}, s_{i \rightarrow i-2} \right), \quad (1)$$

$$f_i^j = \hat{f}_i^j + \mathcal{R} \left( c \left( f_i^{j-1}, \hat{f}_i^j \right) \right), \quad (2)$$

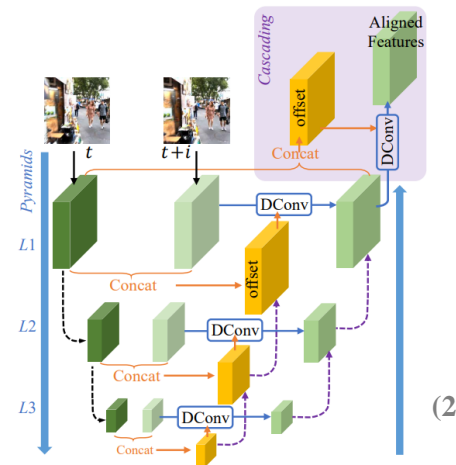
- Grid Propagation은 video sequence의 정보를 취합하는데 적합함
- 기존의 Bidirectional Propagation의 성능을 향상 시키기 위해 Grid Propagation 적용
- Forward, Backward를 번갈아 가며 배치하여 양방향의 정보를 얻음
- 2차 Connection을 추가하여 서로 다른 시공간의 정보 수집

# BasicVSR++: Flow-Guided Deformable Alignment

- Flow based alignment
  - SPyNet or PWCNet을 이용하여 두 frame 간의 optical flow 추출
  - Optical flow를 offset으로 frame  $x_{i\pm 1}$ 을  $x_i$ 에 맞게 warping하여 alignment
- Deformable alignment
  - Deformable Convolution을 이용한 Alignment
  - 두 frame을 Concatenation 하여 Conv를 통과시킨 offset을 이용해 Deformable convolution 진행



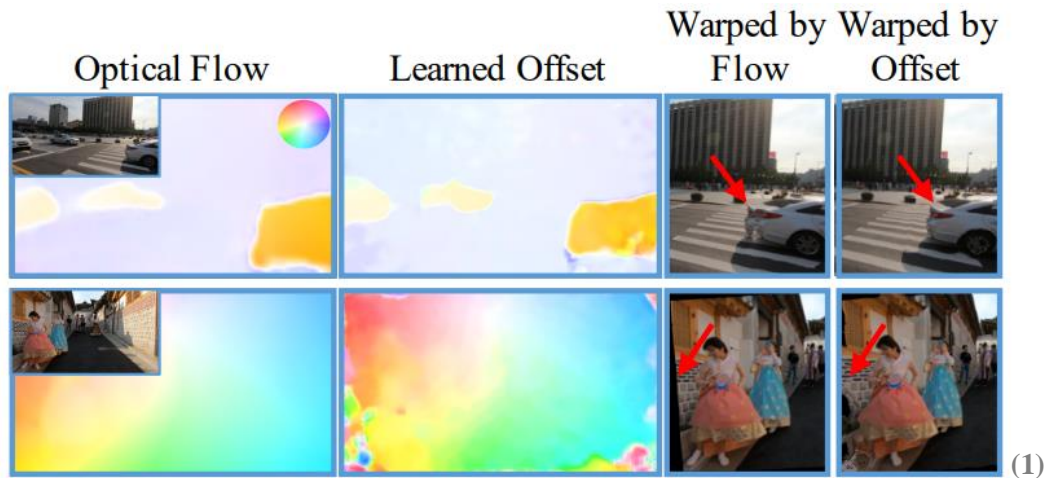
(1)



(2)

# BasicVSR++: Flow-Guided Deformable Alignment

- Deformable Alignment

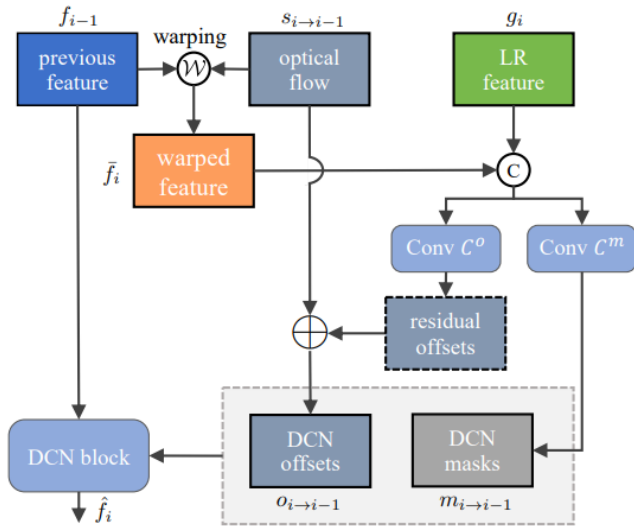


- DCN offset이 Optical Flow와 같은 역할을 함
- Offset diversity와, 학습할 수 있다는 점으로 Flow base로 한 방식보다 정교한 성능
- 하지만 Deformable alignment는 훈련 과정에서 offset overflow 발생 하는 불안정성이 있음



# BasicVSR++: Flow-Guided Deformable Alignment

- Flow-Guided Deformable Alignment



$$\bar{f}_{i-1} = \mathcal{W}(f_{i-1}, s_{i \rightarrow i-1}), \quad (3)$$

$$\begin{aligned} o_{i \rightarrow i-1} &= s_{i \rightarrow i-1} + \mathcal{C}^o(c(g_i, \bar{f}_{i-1})), \\ m_{i \rightarrow i-1} &= \sigma(\mathcal{C}^m(c(g_i, \bar{f}_{i-1}))). \end{aligned} \quad (4)$$

$$\hat{f}_i = \mathcal{D}(f_{i-1}; o_{i \rightarrow i-1}, m_{i \rightarrow i-1}), \quad (5)$$

- Flow base로 사전 정렬하여 offset 학습을 지원
- Optical flow를 더하여 Residual offset만 학습하도록 하여 안정적인 학습 가능
- Deformable alignment의 offset diversity를 유지하고, instability를 해결

# BasicVSR++: Ablation studies

- Ablation studies

Table 3. **Ablation studies.** Each component brings significant improvements in PSNR, verifying their effectiveness.

	(A)	(B)	(C)	BasicVSR++
Flow-Guided Deform. Align.		✓	✓	✓
Second-Order Propagation			✓	✓
Grid Propagation				✓
PSNR (dB)	31.48	31.94	32.08	32.39

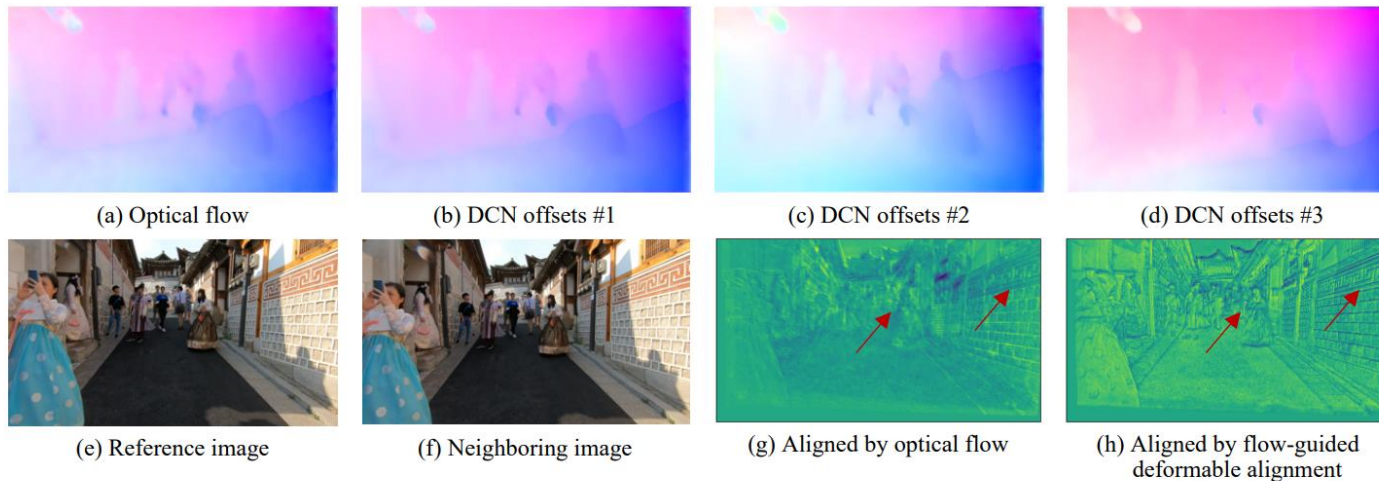


Figure 8. **Benefits of flow-guided deformable alignment.** (a-d) While DCN offsets are highly similar to optical flow, there are noticeable differences. (e-f) The reference and neighboring images. (g) The feature aligned by optical flow, as in BasicVSR, experiences blurry edges. (h) The feature aligned by our proposed module in BasicVSR++ is sharper and preserves more details, as indicated by the red arrows.

# BasicVSR++: Results

Table 1. **Quantitative comparison (PSNR/SSIM)**. All results are calculated on Y-channel except REDS4 [21] (RGB-channel). **Green** and **blue** colors indicate the best and the second-best performance, respectively. The runtime is computed on an LR size of  $180 \times 320$ . A  $4 \times$  upsampling is performed following previous studies. Blank entries correspond to results that are not reported in previous works.

	Params (M)	Runtime (ms)	BI degradation			BD degradation		
			REDS4 [21]	Vimeo-90K-T [35]	Vid4 [19]	UDM10 [37]	Vimeo-90K-T [35]	Vid4 [19]
Bicubic	-	-	26.14/0.7292	31.32/0.8684	23.78/0.6347	28.47/0.8253	31.30/0.8687	21.80/0.5246
VESPCN [1]	-	-	-	-	25.35/0.7557	-	-	-
SPMC [27]	-	-	-	-	25.88/0.7752	-	-	-
TOFlow [35]	-	-	27.98/0.7990	33.08/0.9054	25.89/0.7651	36.26/0.9438	34.62/0.9212	-
FRVSR [24]	5.1	137	-	-	-	37.09/0.9522	35.64/0.9319	26.69/0.8103
DUF [15]	5.8	974	28.63/0.8251	-	-	38.48/0.9605	36.87/0.9447	27.38/0.8329
RBPN [9]	12.2	1507	30.09/0.8590	37.07/0.9435	27.12/0.8180	38.66/0.9596	37.20/0.9458	-
EDVR-M [31]	3.3	118	30.53/0.8699	37.09/0.9446	27.10/0.8186	39.40/0.9663	37.33/0.9484	27.45/0.8406
EDVR [31]	20.6	378	31.09/0.8800	37.61/0.9489	27.35/0.8264	39.89/0.9686	37.81/0.9523	27.85/0.8503
PFNL [37]	3.0	295	29.63/0.8502	36.14/0.9363	26.73/0.8029	38.74/0.9627	-	27.16/0.8355
MuCAN [17]	-	-	30.88/0.8750	37.32/0.9465	-	-	-	-
TGA [13]	5.8	-	-	-	-	-	37.59/0.9516	27.63/0.8423
RLSP [8]	4.2	49	-	-	-	38.48/0.9606	36.49/0.9403	27.48/0.8388
RSDN [12]	6.2	94	-	-	-	39.35/0.9653	37.23/0.9471	27.92/0.8505
RRN [14]	2.4	45	-	-	-	38.06/0.9644	-	27.69/0.8488
BasicVSR [3]	6.3	63	31.42/0.8909	37.18/0.9450	27.24/0.8251	39.96/0.9694	37.53/0.9498	27.96/0.8553
IconVSR [3]	8.7	70	<u>31.67/0.8948</u>	37.47/0.9476	<u>27.39/0.8279</u>	40.03/0.9694	<u>37.84/0.9524</u>	28.04/0.8570
VSR-Trans [2]	32.6	4312	31.06/0.8815	37.71/0.9494	27.36/0.8258	-	-	-
<b>BasicVSR++</b>	7.3	77	<b>32.39/0.9069</b>	<b>37.79/0.9500</b>	<b>27.79/0.8400</b>	<b>40.72/0.9722</b>	<b>38.21/0.9550</b>	<b>29.04/0.8753</b>



Sequence 0216, Clip 024



Bicubic 23.79 dB      RBPN 28.65 dB      EDVR-M 28.06 dB      EDVR 29.64 dB



BasicVSR 28.25 dB      IconVSR 28.79 dB      **BasicVSR++ (ours) 30.80 dB**      GT PSNR

감사합니다.