

Burst Super Resolution

남윤찬

Vision & Display Systems Lab.

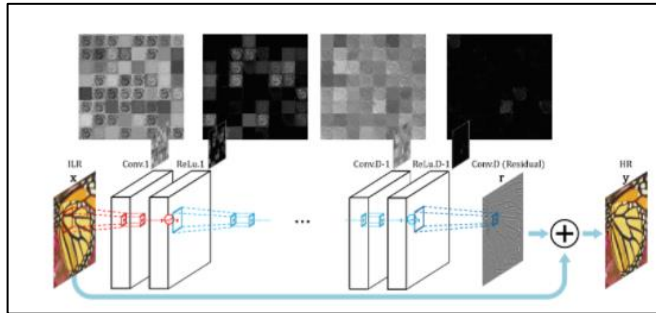
Dept. of Electronic Engineering, Sogang University

개요

- Super Resolution
 - SISR
 - MISR
- Burst Super Resolution
 - Deep Burst Super Resolution (CVPR 2021)
 - Deep Reparametrization of Multi-Frame Super Resolution (ICCV 2021)
- Conclusion

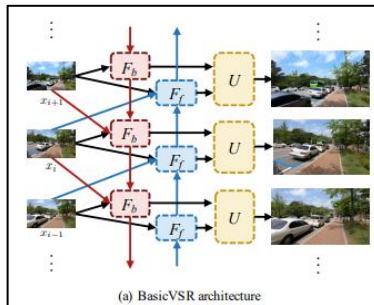
Super Resolution

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

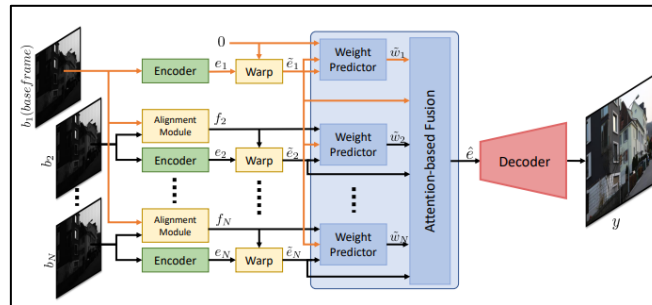


One-to-one (VDSR)

- MISR(Multi-Image Super-Resolution)

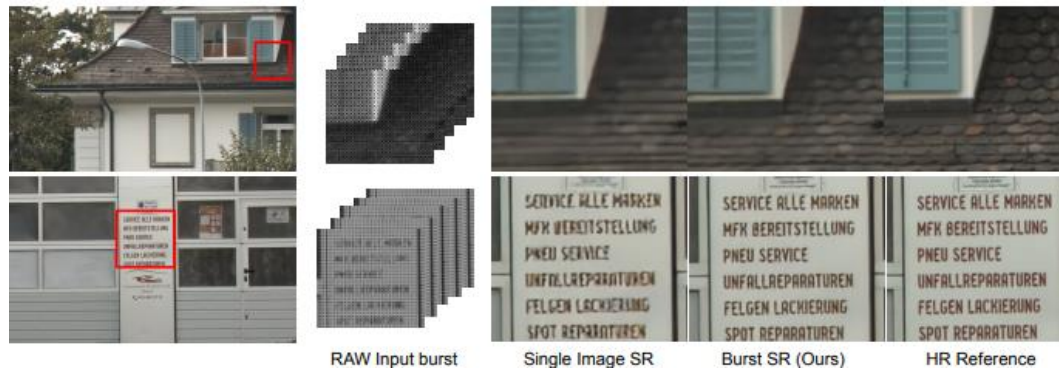


Many-to-many (BasicVSR)



Many-to-one (DBSR)

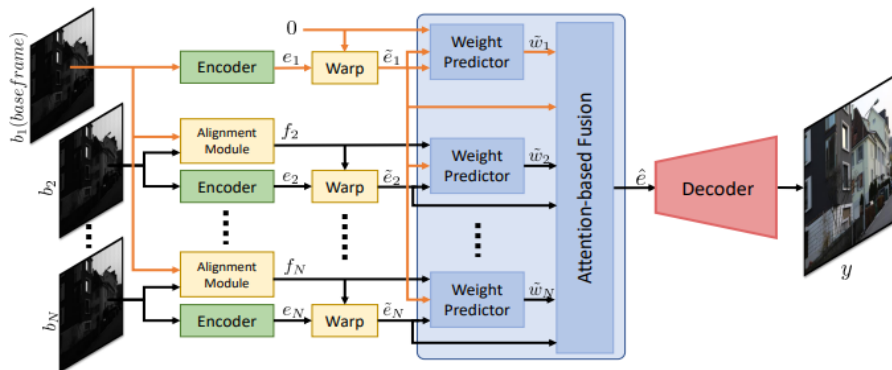
Deep Burst Super Resolution(DBSR)



- Abstract

- Multiple RAW image input, generates a denoised, super-resolved RGB image output
- Input frames들의 deep embeddings를 pixel wise optical flow를 사용하여 alignment.
- Attention-based fusion model을 이용하여 adaptively merge 수행
- Real-world data를 training, evaluation을 수행하기 위해서 BurstSR dataset를 소개

Deep Burst Super Resolution(DBSR)



• Architecture

- N개의 RAW Burst LR image $\{b_i\}_{i=1}^N$ 를 Input으로 받고 b_1 을 reference(base) frame으로 선정
- Input frames을 Encoder를 통해 deep feature representation $\{e_i\}_{i=1}^N$ 생성
- Alignment module 을 통해 b_i 과 $\{b_i\}_{i=2}^N$ 간의 flow estimation(PWCNet)
- 추출한 offsets으로 각 features를 b_1 을 기준으로 warping하여 alignment 수행
- Aligned features 를 제안한 Attention based fusion module 을 이용하여 combine
- Decoder 를 통해 SR output image 출력

Deep Burst Super Resolution(DBSR)

- Encoder

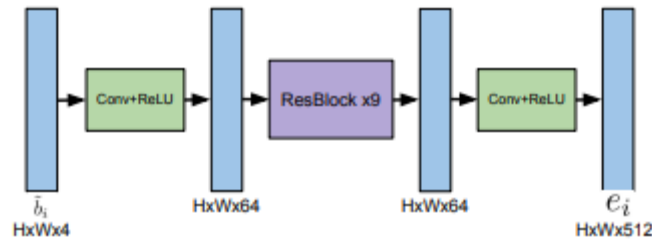
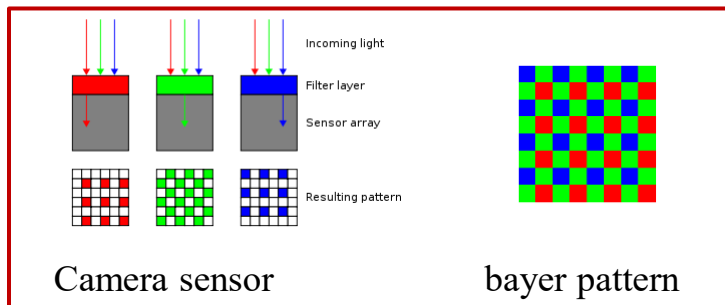


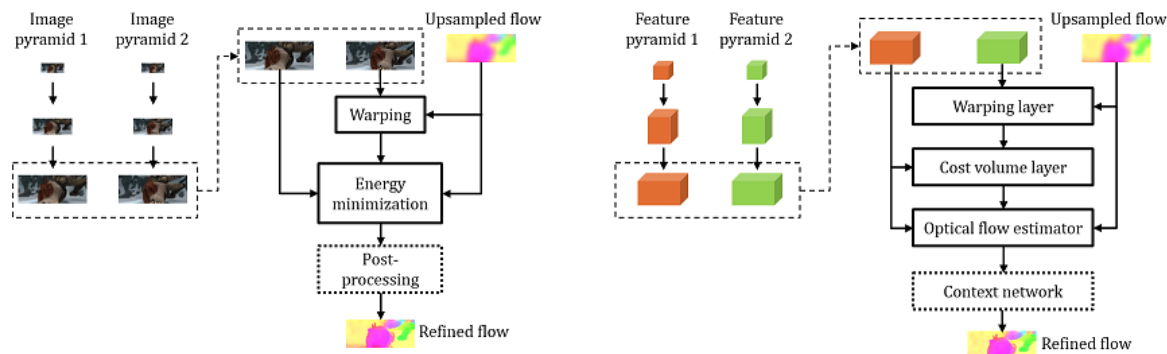
Figure 4. The network architecture employed for the Encoder module E .

- Raw bayer pattern의 2x2 block을 채워 넣어 4-channel image 생성하여 input으로 사용
- 각 burst frame당 $D=512$ 의 feature 생성 ($B \times D \times H \times W$)



Deep Burst Super Resolution(DBSR)

- Alignment module (PWC-Net)

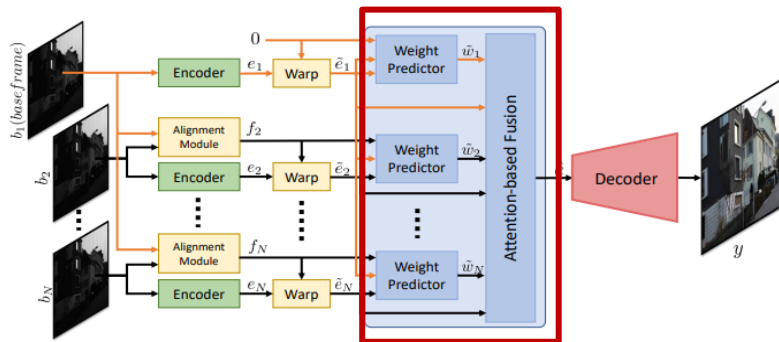


- BurstSR에서 중요한 challenges중 하나는 이미지 간의 pixel-wise displacement를 알 수 없다는 것
- b_1 을 base로 b_i frames들 간의 pixel-wise flow를 계산하고 feature level alignment를 수행

$$\tilde{e}_i = \phi(e_i, f_i), \quad f_i = F(\tilde{b}_i, \tilde{b}_1)$$

- Flow estimation network로 pretrained PWC-Net model을 사용

Deep Burst Super Resolution(DBSR)



• Fusion Module

- Element-wise max or average pool 방법으로 fusion을 취하는 것은 만족스럽지 못한 결과
- Fusion module은 image content, noise level등에 따라 adaptively 하게 적용
- Weight predictor net에 의해 element-wise fusion weight를 예측하고 fusion 하는 Attention based fusion module 제안

Deep Burst Super Resolution(DBSR)

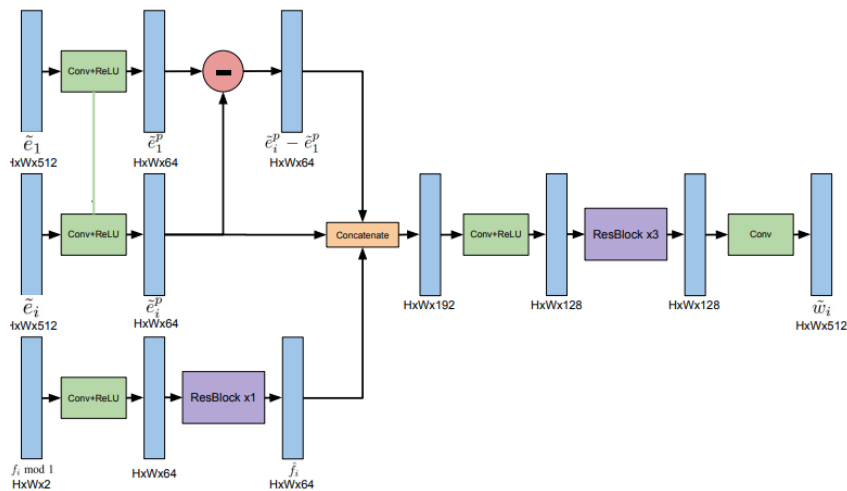


Figure 5. The network architecture employed for the Weight Predictor module W .

• Fusion module (weight predictor)

- Aligned feature map \tilde{e}_i 와 flow vector f_i 를 input으로 사용
- \tilde{e}_i 에서 계산 효율성을 위해 512 에서 64 로 차원 축소 \tilde{e}_1^p , \tilde{e}_i^p
- $\tilde{e}_1^p - \tilde{e}_i^p$ 을 수행하여 residual map r_i 생성
- \tilde{e}_i , \hat{f}_i , r_i 를 concatenate 수행하여 weight \tilde{w}_i 생성 $\tilde{w}_i = W(\tilde{e}_1, r_i, \hat{f}_i)$

Deep Burst Super Resolution(DBSR)

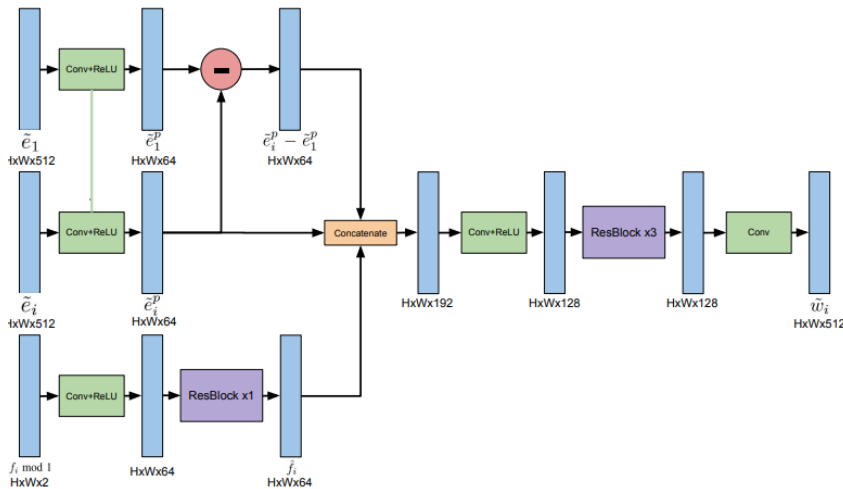


Figure 5. The network architecture employed for the Weight Predictor module W .

- Noisy input frames를 Sub-pixel 정확도로 정렬
- Alignment error에 강하면서도 alignment된 frame의 정보를 효과적으로 fusion
 - Base frame map \tilde{e}_1^p : local image content 정보
 - Residual r_i : alignment error를 추정, misalign된 부분에 낮은 fusion weight를 할당
 - Flow vector f_i : image의 sub-pixel sampling location 정보

Deep Burst Super Resolution(DBSR)

- Fusion module

- weight \tilde{w}_i 를 softmax를 취해 정규화 한 후 최종 attention weights를 생성, aligned feature와 element-wise multiplication을 하여 더해준다.

$$\hat{e} = \sum_{i=1}^N w_i \cdot e_i, \quad w_i = \frac{e^{\tilde{w}_i}}{\sum_j e^{\tilde{w}_j}}$$

Deep Burst Super Resolution(DBSR)

- Decoder

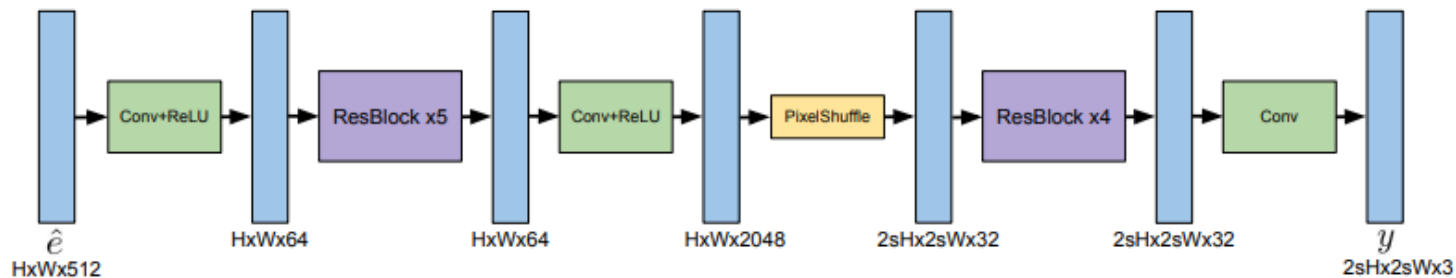
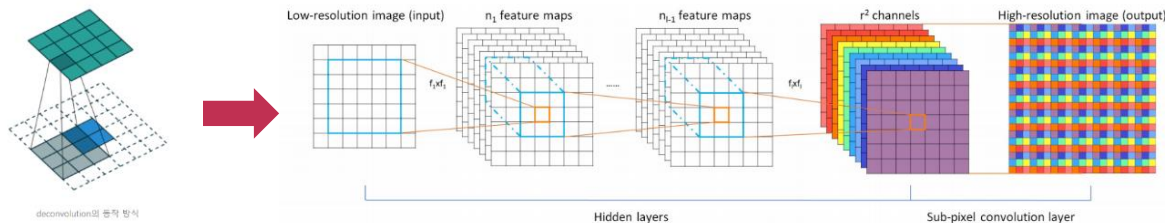


Figure 6. The network architecture employed for the Decoder module D .

- Sub-pixel convolution을 이용하여 upsampling



- LR 이미지의 feature map을 r^2 만큼 채널 수를 늘려준 다음 그 feature map을 순서대로 조합하여 $H * W * C * r^2$ tensor를 $rH * rW * C$ 의 고해상도 이미지로 변환

Deep Burst Super Resolution(DBSR)

- Experiments & Result

	Synthetic data			BurstSR		
	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
Single Image	36.42	0.123	0.913	46.41	0.041	0.979
Burst-2	34.90	0.133	0.893	46.10	0.040	0.977
Burst-4	37.18	0.092	0.927	47.06	0.033	0.981
Burst-8	38.61	0.084	0.941	47.52	0.031	0.983
Burst-14	39.09	0.084	0.945	47.76	0.030	0.984

Table 1. Comparison of the baseline SISR network with our multi-frame approach, evaluated using different number of input frames.

	Synthetic data			BurstSR		
	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
Ours	38.61	0.084	0.941	47.52	0.031	0.983
No Alignment	36.66	0.119	0.915	46.50	0.040	0.979
Single Image	36.42	0.123	0.913	46.41	0.041	0.979

Table 2. Comparison of our approach performing explicit alignment with a baseline which does not employ an alignment module.

	Synthetic data			BurstSR		
	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
Ours	38.61	0.084	0.941	47.52	0.031	0.983
MaxPool	36.24	0.116	0.912	46.74	0.039	0.980
AvgPool	35.45	0.131	0.902	46.53	0.040	0.979
Concatenate	37.80	0.098	0.928	47.17	0.034	0.981
RecMerge	37.55	0.098	0.927	47.12	0.033	0.981

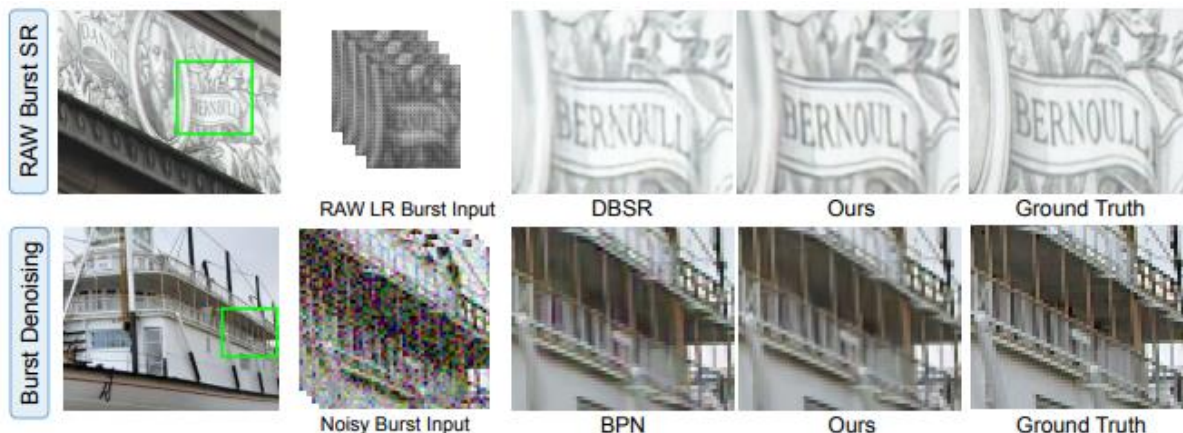
Table 3. Analysis of different fusion approaches for merging the information from input frames.

	Synthetic data			BurstSR		
	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow
Only Feature	37.46	0.101	0.927	47.11	0.034	0.981
Only Residual	38.14	0.093	0.935	47.46	0.031	0.982
Residual+Base	38.41	0.085	0.939	47.46	0.030	0.983
Residual+Base+Flow	38.61	0.084	0.941	47.52	0.031	0.983

Table 4. Impact of different inputs used by the weight predictor.



Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)



- Abstract

- Multi-frame image restoration task에서 흔히 사용되는 maximum a posteriori formulation(MAP) 기법의 deep reparametrization을 제안
- Deep reparametrization을 통해 latent 공간의 image 형성 과정을 직접 모델링 하고 학습된 image priors를 예측에 통합시킬 수 있다.
- Deep learning을 장점을 활용하는 동시에 MAP가 가지는 multi-frame fusion의 이점도 이용

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)

- Method
 - 기존의 multi-frame image restoration problem

$$x_i = H(\phi_{m_i}(y)) + \eta_i \quad (2)$$

$$\hat{y} = \arg \min_y \sum_{i=1}^N \|x_i - H(\phi_{m_i}(y))\|_2^2 + \mathcal{R}(y) \quad (3)$$

- x_i 와 $\bar{x}_i = H(\phi_{m_i}(y))$ 간의 error를 최소화 하는 방법으로 MAP estimation을 이용한다.
- $R(y)$: 원본 이미지 y 의 prior knowledge를 통합하는 regularization term

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)

$$\hat{y} = \arg \min_y \sum_{i=1}^N \|x_i - H(\phi_{m_i}(y))\|_2^2 + \mathcal{R}(y) \quad (3)$$

- multiple frames information을 intergrating 하는데 주된 방식
- degradation operator H와 regularizer R을 수동적으로 설정해줘야 하는 문제
- 복잡한 noise 분포를 generalize 하는 유연성이 부족함
- Deep Reparametrization을 제안

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)

<Step 1>

- 기존의 L2 distance $\|x_i - \bar{x}_i\|_2$ 를 최소화 하는 문제를 encoder E 를 통과하여 얻은 embeddings들 간의 L2 distance $d(x_i, \bar{x}_i) = \|E(x_i) - E(\bar{x}_i)\|_2$ 를 최소화 하는 것으로 변경
 - 인코더 E의 depth와 비 선형성 때문에 복잡한 noise와 error 분포에 좀 더 적합하고 매우 유연한 metric 생성

<Step 2>

- 이미지 y의 latent deep representation ‘z’를 생성하여 decoder network ‘D’를 이용하여 y를 대신한다. ($y = D(z)$)
 - z는 y의 직접 매개변수이기 때문에 z에 대한 objective를 최적화 하고 최종 latent representation \hat{z} 을 구하면 최종 이미지 $\hat{y} = D(\hat{z})$ 를 구할 수 있다.
 - 출력 이미지 y대신 latent representation z에 대해 최적화 하기 때문에 decoder D는 preconditioner 역할을 하여 더 빠르게 수렴

$$L(z) = \sum_{i=1}^N \|E(x_i) - E \circ H \circ \phi_{m_i} \circ D(z)\|_2^2 + \mathcal{R}(D(z))$$
$$\hat{y} = D(\hat{z}), \quad \hat{z} = \arg \min L(z). \quad (4)$$

$$L(z) = \sum_{i=1}^N \|E(x_i) - \underbrace{E \circ H \circ D}_{G} \circ \phi_{m_i}(z)\|_2^2 + \mathcal{R}(D(z)), \quad (5)$$

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)

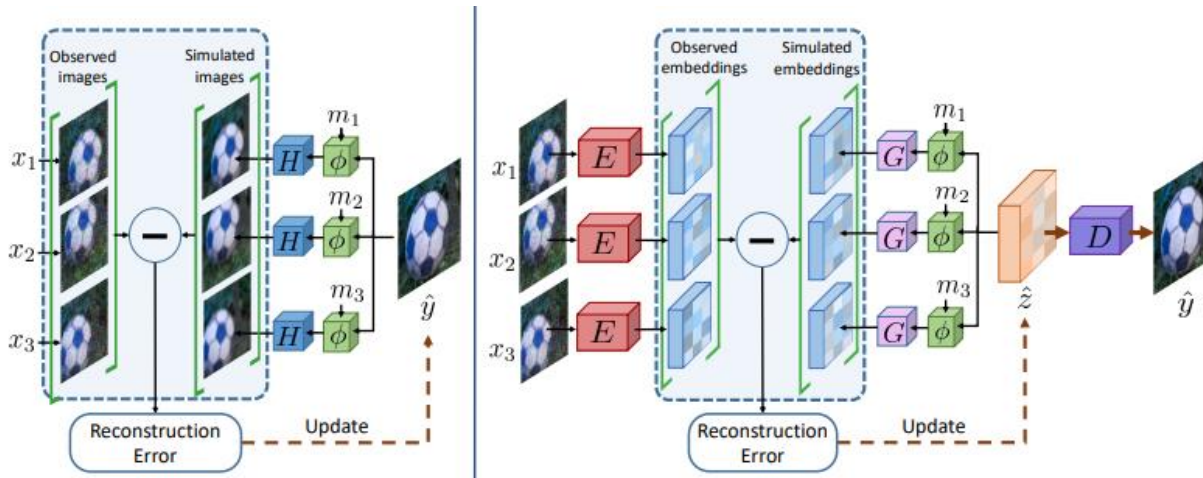
<Step 3>

- Degradation operator H 는 일반적으로 unknown이며 추정 또는 학습이 필요
 - H 를 별도의 신경망으로 만들어 매개변수화 시켜 학습 시킬 수 있음
 - $E \circ H \circ D$ 를 deep feature space degradation operator G 로 대체
 - H 를 따로 학습하는 대신 operator G 로 통합하여 시뮬레이션 된 embedding $G(\phi_{m_i}(z))$ 를 얻는데 사용

$$L(z) = \sum_{i=1}^N \|E(x_i) - \underbrace{E \circ H \circ D}_G \circ \phi_{m_i}(z)\|_2^2 + \mathcal{R}(D(z)), \quad \rightarrow \quad L(z) = \sum_{i=1}^N \|E(x_i) - G(\phi_{m_i}(z))\|_2^2 + \mathcal{Q}(z). \quad (6)$$

(5)

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)



$$\hat{y} = \arg \min_y \sum_{i=1}^N \|x_i - H(\phi_{m_i}(y))\|_2^2 + \mathcal{R}(y)$$

$$\|x_i - \bar{x}_i\|_2 \rightarrow \begin{aligned} d(x_i, \bar{x}_i) &= \|E(x_i) - E(\bar{x}_i)\|_2 \\ &= \|H(\phi_{m_i}(y)) - H(\phi_{m_i}(\bar{y}))\|_2 \\ &= \|G(\phi_{m_i}(z)) - G(\phi_{m_i}(\bar{z}))\|_2 \\ &= \|G(\phi_{m_i}(z)) - G(\phi_{m_i}(\bar{z}))\|_2 \end{aligned}$$



$$L(z) = \sum_{i=1}^N \|E(x_i) - E \circ H \circ \phi_{m_i} \circ D(z)\|_2^2 + \mathcal{R}(D(z))$$

$$\hat{y} = D(\hat{z}), \quad \hat{z} = \arg \min L(z). \quad (4)$$

$$L(z) = \sum_{i=1}^N \|E(x_i) - \underbrace{E \circ H \circ D}_{G} \circ \phi_{m_i}(z)\|_2^2 + \mathcal{R}(D(z)), \quad (5)$$

$$L(z) = \sum_{i=1}^N \|E(x_i) - G(\phi_{m_i}(z))\|_2^2 + \mathcal{Q}(z). \quad (6)$$

• 기준

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)

- Certainty Predictor

$$L(z) = \sum_{i=1}^N \|E(x_i) - G(\phi_{m_i}(z))\|_2^2 + \mathcal{Q}(z). \quad (6)$$

- 각 frame 마다 다른 노이즈, 추론된 m_i 의 오류 등을 고려하여 certainty predictor W 추가

$$L(z) = \sum_{i=1}^N \|v_i \cdot (E(x_i) - G(\phi_{m_i}(z)))\|_2^2 + \lambda \|z\|_2^2$$

where $v_i = W(\{E(x_j)\}_{j=1}^N, m_i, n_i)$. (7)

- W 를 통해 계산된 weights v_i 를 frame wise하게 적용
- W 는 DBSR의 weight predictor와 동일

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)

- Optimization

- Steepest descent algorithm (\equiv gradient descent)
- Gradient : $g^j = \nabla L(z^j)$
- Step size (learning rate) : $\alpha^j = \operatorname{argmin}_{\alpha} L(z^j - \alpha g^j)$
- $z^{j+1} = z^j - \alpha^j g^j$

$$g^j = -2 \sum_{i=1}^N \phi_{m_i}^T G^* \cdot (v_i^2 \cdot (E(x_i) - G * \phi_{m_i}(z^j))) + 2\lambda z^j$$

$$\alpha^j = \frac{\|g^j\|_2^2}{\sum_{i=1}^N 2\|v_i \cdot (G * \phi_{m_i}(g^i))\|_2^2 + 2\lambda \|g^j\|_2^2} \quad (8)$$

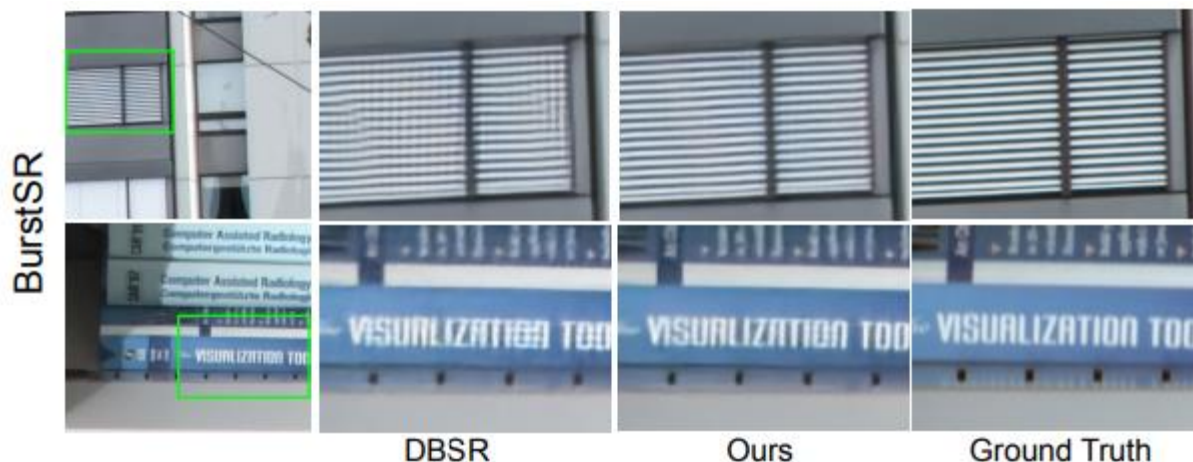
$$z^{j+1} = z^j - \alpha^j g^j .$$

- Initializer P

- 수렴 속도를 향상시키기 위해 초기 latent representation $z^0 = P(E(x_i))$ 로 설정

Deep Reparametrization of Multi-Frame Super Resolution (DeepRep)

- Experiments & Result



	SyntheticBurst			BurstSR			Time (s)
	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	
SingleImage	36.86	0.113	0.919	46.60	0.039	0.979	0.02
HighResNet [12]	37.45	0.106	0.924	46.64	0.038	0.980	0.11
DBSR [2]	40.76	0.053	0.959	48.05	0.025	0.984	0.24
Ours	41.56	0.045	0.964	48.33	0.023	0.985	0.40

Table 1. Comparison on the SyntheticBurst and real-world BurstSR validation dataset from [2].

Conclusion

- Real-world Burst SR에 대한 기법 소개
 - Deep Burst Super Resolution (CVPR 2021)
 - Deep Reparametrization of Multi-Frame Super Resolution (ICCV 2021)
- Muti-frame super resolution의 critical한 부분인 alignment 부분 개선
 - DCN alignment
 - Other flow estimation model
- DeepRep model의 MAP기법을 이용한 image fusion model 개선
 - Parameter 수 감소
 - 성능 향상