Ada-VSR: Adaptive Video Super-Resolution with Meta-Learning

Akash Gupta, Padmaja Jonnalagedda, Bir Bhanu, Amit K. Roy-Chowdhury

University of California, Riverside CA

ACM multimedia
Chengdu, China OCT 20-24 2021

Adaptive Video Super-Resolution

A meta-transfer learning framework for the task of blind spatio-temporal video super-resolution.

Key Innovations:

- ✓ Meta-learning framework **Ada-VSR** for the task of joint spatio-temporal super-resolution by leveraging external and internal learning.
- ✓ External learning to learn weights that can easily adapt to novel conditions for super-resolution tasks.
- ✓ **Ada-VSR** reduces the computational time by reducing the gradient steps required during internal learning.

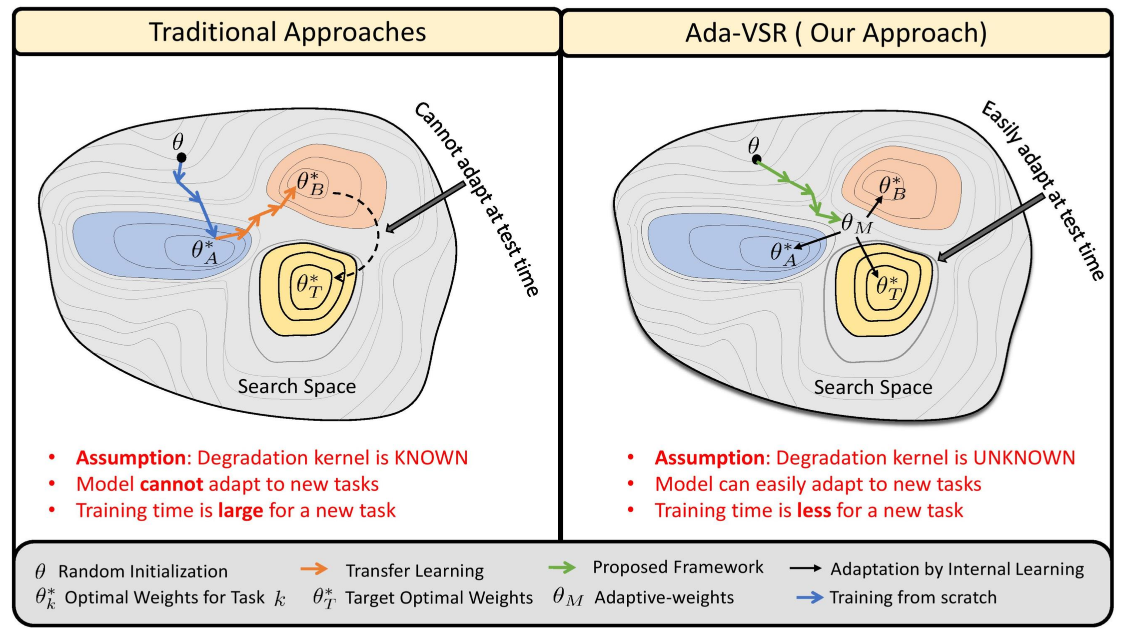
Problem Statement

How can we make our network adapt quickly for blind settings?

- Model trained with the assumption of **known degradation kernel** works well for a video with similar degradation.
- Deep internal learning is used to tackle the blind settings but it has **large computational time**.
- **Meta-learning** leverages an external dataset to learn parameters that can **quickly adapt to blind settings** during test using internal statistics of the input test video.

Conceptual Overview

Learning weights that can adapt quickly to novel conditions!



Meta-Transfer Learning

Input : High-resolution high-frame rate dataset $\mathcal{D}_{HR}$ and task distribution $p(\mathcal{T})$

Input : α, β task-specific and adaptation learning rate

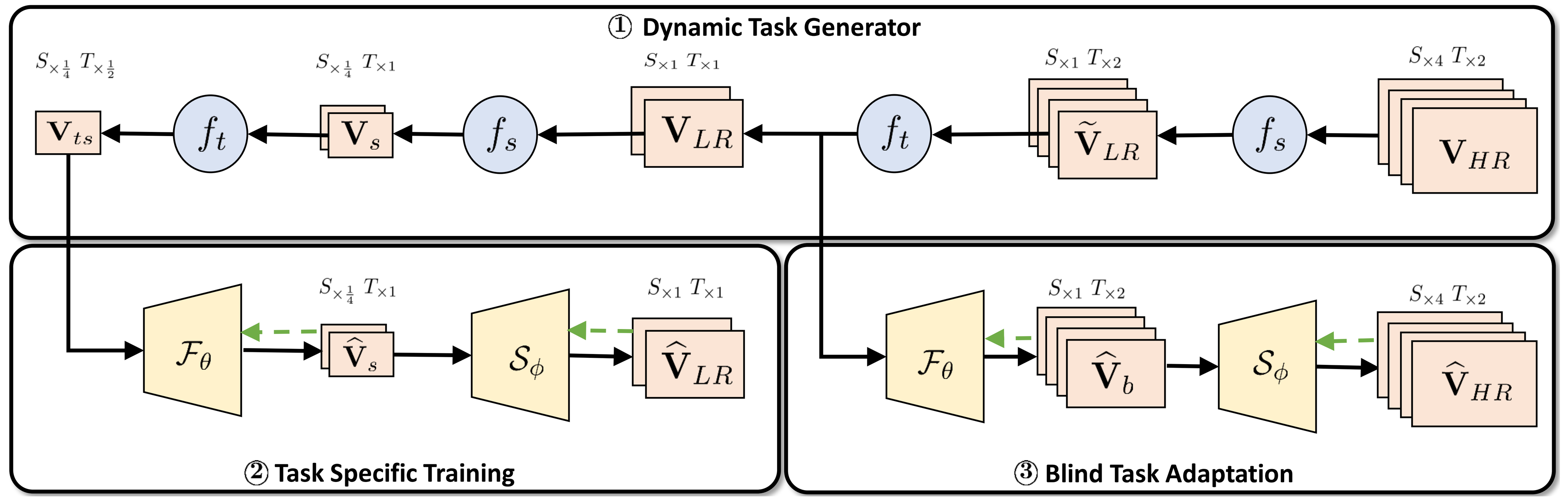
Output : **Ada-VSR** model parameters θ and ϕ

$$\widehat{\mathbf{V}}_s = \mathcal{F}_\theta(\mathbf{V}_{ts}) \qquad \widehat{\mathbf{V}}_{LR} = \mathcal{S}_\phi(\widehat{\mathbf{V}}_s) \tag{1}$$

$$\mathcal{L}^{tr}_{\mathcal{T}_i} = \sum_{\mathcal{D}_{tr}}\sum_{n_i}\Big(\mathcal{L}\big(\widehat{\mathbf{V}}_s, \mathbf{V}_s\big) + \mathcal{L}\big(\widehat{\mathbf{V}}_{LR}, \mathbf{V}_{LR}\big)\Big) \tag{2}$$

$$\mathcal{L}^{te}_{\mathcal{T}_j}(\theta_i, \phi_i) = \mathcal{L}^{te}_{\mathcal{T}_j}(\theta - \alpha\nabla_\theta\mathcal{L}^{tr}_{\mathcal{T}_i}, \quad \phi - \alpha\nabla_\phi\mathcal{L}^{tr}_{\mathcal{T}_i}) \tag{3}$$

```
Sample task batch D_tr, D_te for the task T_i, T_j ~ p(T)
/* Task-Specific Training (inner loop) */
for all T_i do
    Compute meta-training loss (D_tr): L^tr_{T_i}(θ, φ)
    Adapt parameters with gradient descent:
    θ_i = θ − α∇_θ L^tr_{T_i}(θ, φ),   φ_i = φ − β∇_φ L^tr_{T_i}(θ, φ)
/* Blind Task Adaptation (outer loop) */
Update θ and φ with respect to average test loss (D_te):
θ ← θ − α∇_θ Σ_{T_i~p(T)} L^te_{T_j}(θ_i, φ_i)
φ ← φ − β∇_φ Σ_{T_i~p(T)} L^te_{T_j}(θ_i, φ_i)
```



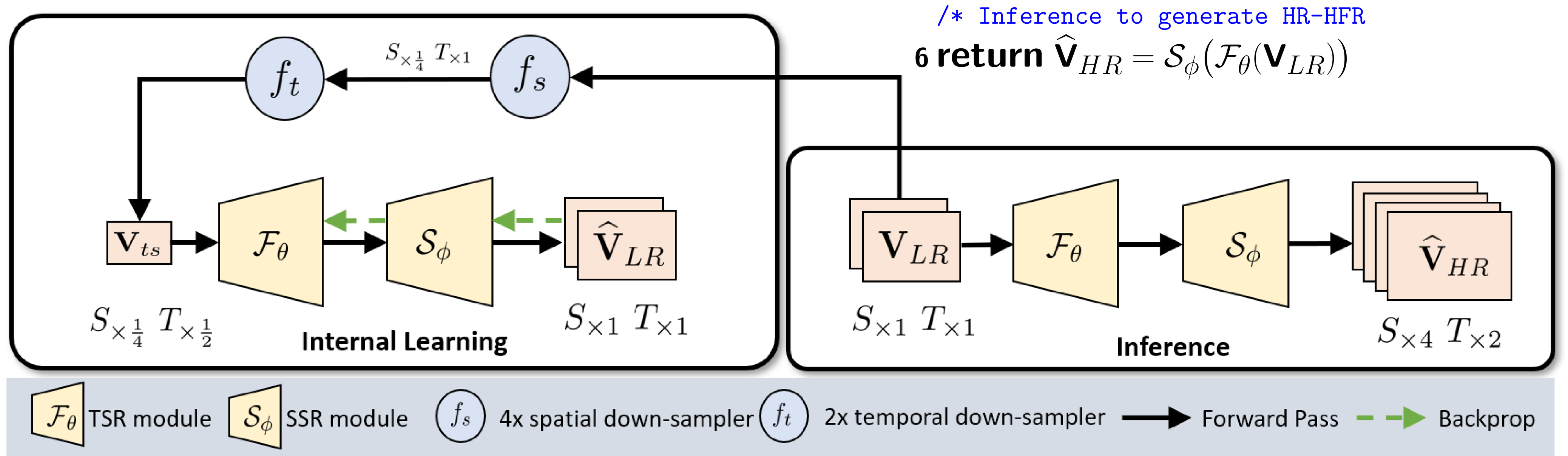
Internal Learning and Inference

Input : LR-LFR test video $\mathbf{V}_{LR}$, meta-transfer trained model parameter θ, ϕ, number of gradient updates n and learning rate γ

Output : High-resolution high-frame rate video $\widehat{\mathbf{V}}_{HR}$

$$\mathcal{L}_{int} = \Big\|\mathcal{S}_\phi\big(\mathcal{F}_\theta(\mathbf{V}_{ts})\big) - \mathbf{V}_{LR}\Big\|_1 \tag{4}$$

```
Generate down-sampled video V_ts by down-sampling
  V_LR with corresponding blur kernel.
/* Internal Learning */
for n steps do
    Evaluate loss L_int(θ, φ) using (4)
    Update θ ← θ − γ∇_θ L_int(θ, φ)
    Update φ ← φ − γ∇_φ L_int(θ, φ)
/* Inference to generate HR-HFR */
return V̂_HR = S_φ(F_θ(V_LR))
```



Quantitative Results on Vimeo Dataset

Method	Vimeo-90K Slow			Vimeo-90K Medium			Vimeo-90K Fast		
End-to-end Framework	PSNR ↑	SSIM ↑	NIQE ↓	PSNR ↑	SSIM ↑	NIQE ↓	PSNR ↑	SSIM ↑	NIQE ↓
Zooming Slow-Mo [1]	33.29	0.91	6.94	35.24	0.93	7.35	36.43	0.93	8.41
Temporal Profile [2]	**33.40**	**0.92**	6.17	35.55	0.94	6.37	36.29	0.93	7.13
Ada-VSR (Ours)	33.36	**0.92**	**6.12**	**35.91**	**0.95**	**6.33**	**36.52**	**0.95**	**6.99**

Quantitative Analysis on Vid4 Dataset

Method	Vid4		
End-to-end Framework	PSNR ↑	SSIM ↑	NIQE ↓
Zooming Slow-Mo [1]	26.30	0.80	5.62
Temporal Profile [2]	26.50	0.82	5.48
Ada-VSR (Ours)	**26.98**	**0.84**	**5.40**

External-Training		Vid4 [3]		
Spatial	Temporal	PSNR ↑	SSIM↑	NIQE↓
✓	✗	25.98	0.80	5.77
✗	✓	26.27	0.81	5.59
✓	✓	**26.98**	**0.84**	**5.40**

Qualitative Results

Conclusions

- We present an Adaptive Video Super Resolution framework (**Ada-VSR**) to generate high resolution high frame-rate videos from low resolution low frame-rate input videos.
- We leverage external as well as internal learning for spatio-temporal super-resolution.
- The proposed approach is able to achieve superior enhancement while adapting to unknown degradation models as shown in our experiments.

Acknowledgement: The work was partially supported by NSF grants 1664172, 1724341 and 1911197.