DVDnet: A Fast Network for Deep Video Denoising

M. Tassano[1][2], J. Delon[1], T. Veit[2]

[1] MAP5, Université Paris Descartes

[2] GoPro France

IEEE ICIP - Set 2019

Introduction

Video Denoising

Video Denoising

- Compared to image denoising, video denoising is an underexplored domain
- The state-of-the-art in video denoising is mostly defined by patch-based methods
  - E.g.
    - V-BM4D[3], extension of BM3D[1]
    - Video non-local Bayes (VNLB [4]), extension of non-local Bayes[2]
- There are almost no CNN approaches

Video Denoising versus Image Denoising

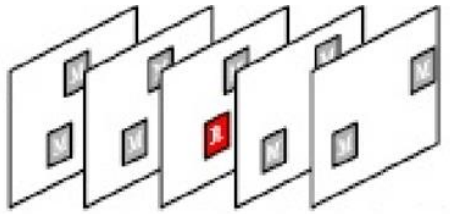
- Compared to image denoising, video denoising is an underexplored domain
  - Video restauration needs to impose a strong temporal coherence in its results

- **Temporal coherence is crucial in the perceived quality of video results [5, 6]**

Introduction: Temporal Coherence

How do video algorithms enforce temporal coherence?

- Two different strategies
  - Motion compensation, e.g. [3, 4, 7]
  - Extend search area to a search volume



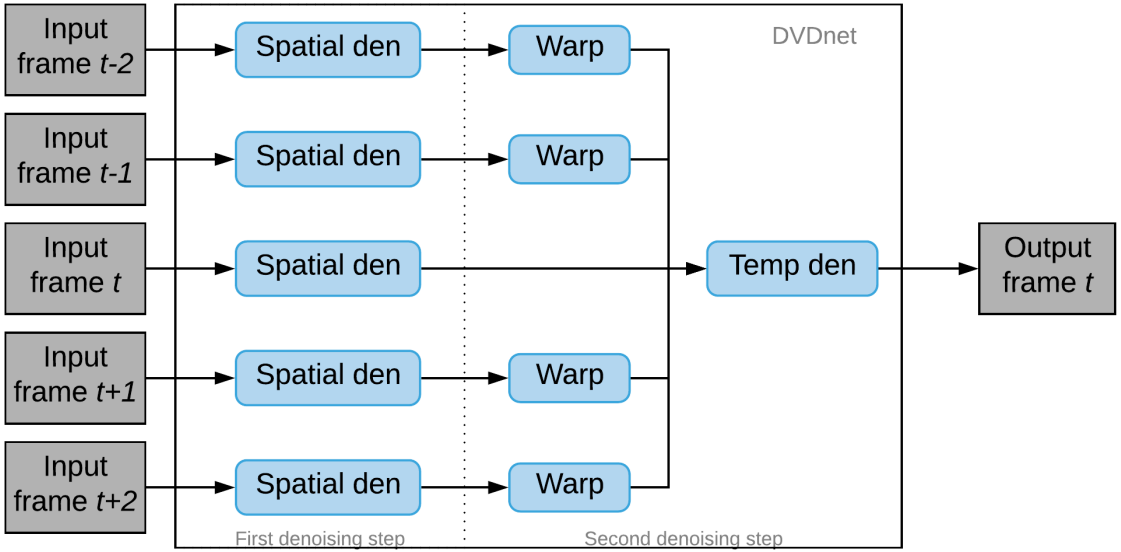
Layout of presentation

- Architecture
- Results
- Running times
- Conclusions

Architecture

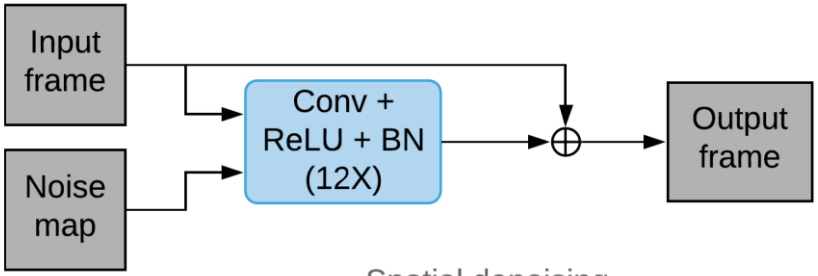
Architecture of DVDnet

Our architecture incorporates these two strategies to enforce temporal coherence

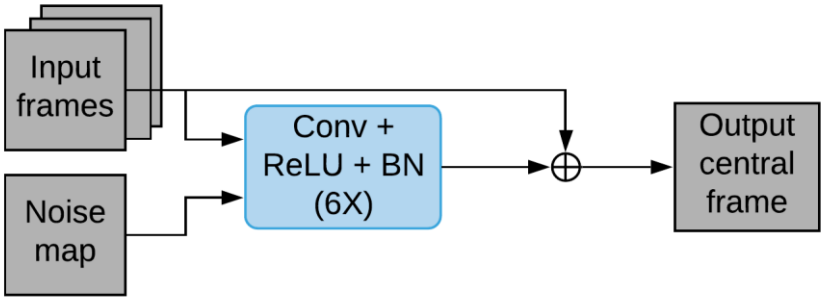


Architecture: spatial and temporal blocks

Two different feed-forward CNN blocks



Spatial denoising



Temporal denoising

Results

Results: versus patch-based methods

DVDnet outperforms other methods for strong noise

	DVDnet	VNLB	V-BM4D
$\sigma = 10$	38.13	**38.85**	37.58
$\sigma = 20$	**35.70**	35.68	33.88
$\sigma = 30$	**34.08**	33.73	31.65
$\sigma = 40$	**32.86**	32.32	30.05
$\sigma = 50$	**31.85**	31.13	28.80

Table: PSNRs on DAVIS testset[8].

Results: versus other CNN methods

Two very recent CNN approaches: VNLnet[9] and FastDVDnet[10]

	V-BM4D	VNLB	VNLnet	FastDVDnet	DVDnet
$\sigma = 10$	37.58	38.85	35.83	38.97	38.13
$\sigma = 20$	33.88	35.68	34.49	35.86	35.70
$\sigma = 30$	31.65	33.73	-	34.06	34.08
$\sigma = 40$	30.05	32.32	32.32	32.80	32.86
$\sigma = 50$	28.80	31.13	31.43	31.83	31.85

Table: PSNRs on the DAVIS testset. Best results are shown in blue, second best in red.

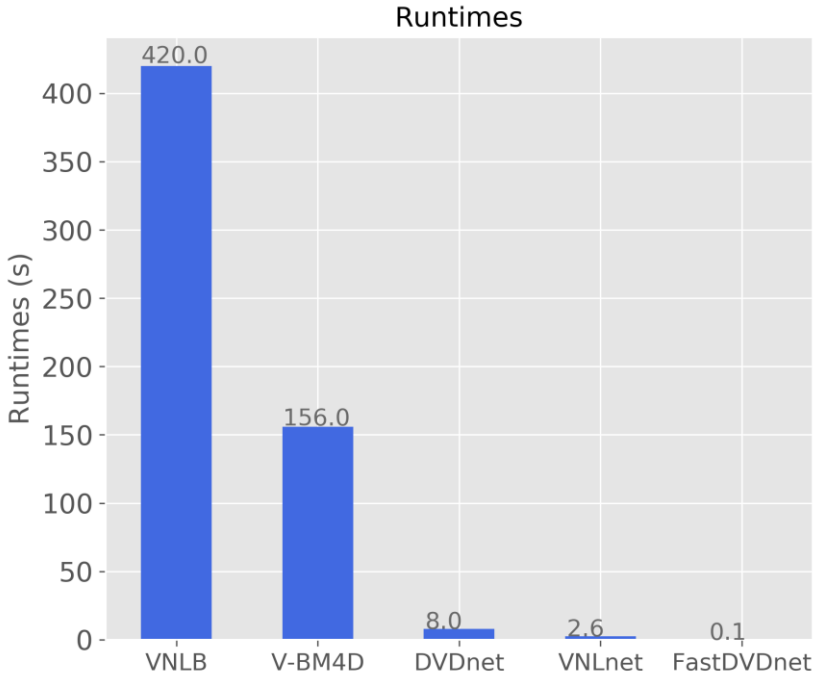
Results: versus V-BM4D

Results: versus VNLB

Running Times

Running Times

- DVDnet is much faster than other patch-based approaches
- Runtimes are comparable to other CNN-based methods



Time to denoise a 960 x 540 color frame

Conclusions

DVDnet improves the state-of-the-art in video denoising

- It’s the best-performing algorithm for medium and strong noise on all tested data
- Performance
  - Remarkable temporal coherence
  - Strong preservation of detail
- Runtimes
  - Much faster than patch-based methods
  - Comparable to other CNN-based approaches
- No hand-tuned parameters involved

References

- [1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3d transform- domain collaborative filtering. IEEE Transactions on image processing, 16(82):3736–3745, 2007.
- [2] Arias, P., & Morel, J.-M. (2015). Towards a Bayesian video denoising method. In S. Battiato, J. Blanc-Talon, G. Gallo, W. Philips, D. Popescu, & P. Scheunders (Eds.), *Computers & Graphics* (Vol. 26, pp. 107–117). Cham: Springer International Publishing.
- [3] Maggioni, M., Boracchi, G., Foi, A., Egiazarian K., “Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms,” IEEE Trans. IP, vol. 21, no. 9, pp. 3952–3966, 2012.
- [4] Arias P. and Morel, J.-M., “Video denoising via empirical Bayesian estimation of space-time patches,” Journal of Mathematical Imaging and Vision, vol. 60, no. 1, pp. 70—93, 2018.
- [5] Seshadrinathan K., Bovik, A. “Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos”. IEEE Transactions on Image Processing, 19(2):335–350, 2010
- [6] Seybold, T. “Noise Characteristics and Noise Perception”. Chapter in Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends, pages 235–265. Springer International Publishing, 2018.
- [7] Buades, A., Lisani, J. “Patch-Based Video Denoising With Optical Flow Estimation”. IEEE Transactions on Image Processing, 25(6):2573–2586, 2016.
- [8] Khoreva A., Rohrbach A., Schiele B. “Video Object Segmentation with Language Referring Expressions”. In Asian Conference on Computer Vision, pages 123–141, Cham, 2019. Springer International Publishing
- [9] Davy A.,Ehret T., Facciolo G., Morel J.-M., Arias P. “Non-Local Video Denoising by CNN”. In IEEE ICIP 2019
- [10] Tassano M., Delon J., Veit T. “FastDVDnet: Towards Real-Time Video Denoising Without Explicit Motion Estimation”. arXiv preprint arXiv:1907.01361, 2019

Thank you! Questions?

For more examples and code visit

https://github.com/m-tassano/dvdnet

https://github.com/m-tassano/fastdvdnet


GoPro
THANKS

Results: DAVIS and Set8

DAVIS	VNLB	V-BM4D	VNLnet	DVDnet	FastDVDnet
$\sigma = 10$	38.85	37.58	35.83	38.13	38.97
$\sigma = 20$	35.68	33.88	34.49	35.70	35.86
$\sigma = 30$	33.73	31.65	-	34.08	34.06
$\sigma = 40$	32.32	30.05	32.32	32.86	32.80
$\sigma = 50$	31.13	28.80	31.43	31.85	31.83
Set8	**VNLB**	**V-BM4D**	**VNLnet**	**DVDnet**	**FastDVDnet**
$\sigma = 10$	37.26	36.05	37.10	36.08	36.43
$\sigma = 20$	33.72	32.19	33.88	33.49	33.37
$\sigma = 30$	31.74	30.00	-	31.79	31.60
$\sigma = 40$	30.39	28.48	30.55	30.55	30.37
$\sigma = 50$	29.24	27.33	29.47	29.56	29.42

Table: PSNRs on the DAVIS and Set8 testsets. Best results are shown in blue, second best in red.

Set8 is composed of 4 GoPro sequences and 4 sequences from the DERF dataset