Supplementary material

This supplementary material presents additional qualitative and quantiative results. In Figure A1 we present additional visual comparison with two burst denoising methods on real images. In Table A1 we present additional experiments on RGB images. Figures A2 and A3 are devoted to super-resolution experiments from real raw data from different smartphones (Google Pixel 3a and 4a, Samsung S7 and S10) and cameras (Panasonic Lumix GX9 and Canon Powershot G7X) and comparison with additional baselines. In Figures A4, we present extreme upsampling results by using synthetic RGB image bursts. In Figure A5, we present restoration results obtained from real images with very low SNR to illustrate the efficiency of our method to perform blind denoising. In Figures A6 and A7, we study the effect of the number of frames in the burst on the reconstruction, both in the low SNR and high SNR settings. Finally, we present failure cases in Figure A8, where fast moving objects are present in the scene.

Comparison with burst denoising methods. We perform additional qualitative comparison on a real image with two burst denoising methods. We compare our method with [10] which performs joint denoising and demosaicking on a burst of raw images. We use the code and the pretrained model made available online. We also use the code and pretrained model of [46]. However the model is only designed to perform grayscale burst denoising, so we perform denoising independently on each RGB channel and then perform demosaicking to get an RGB image. Despite our best efforts for tuning the parameters of these methods to maximize visual quality, the results obtained are not as good as our method (see Figure A1 below). We believe this is not surprising since each one of these methods only addresses a subset of our problem. Adapting them successfully to our general setting is not trivial.

ISP | BKPN [46]+demosaic | Mosa2mosa [10] | Mosa2mosa [10] + EDSR | Ours

Figure A1: Comparison with joint denoising and demosaicking methods.

Evaluation on RGB images. We compare our approach on the BSD68 dataset against state-of-the-art single-image and video super-resolution algorithms (considering a burst as a video sequence) and report the HR image reconstruction accuracy in terms of average PSNR in Table A1. For the training with RGB data, we perform 80 000 iterations of the ADAM optimizer with a batch size of 10, a burst size of 14 and with a learning rate of 3×10^{-5} decaying by a factor 2 after 40 000 iterations. For evaluating the model VSR-DUF [21], we use the code and the pretrained models made available online by the authors. Other single-image reconstruction results are from [48]. In the present setting, we consistently outperform other baselines, notably demonstrating that burst SR cannot simply be addressed effectively by current video SR approaches. We also note that our models perfom better with less blurring (and more aliasing). Finally, we evaluate variations of our model in the same table, notably comparing the registration accuracy achieved by these variants by using the geometrical error presented in [36]. More precisely, we perform a small ablation study by introducing a simpler baseline that does not perform joint alignement and only exploits the coarse registration module (no refine baseline). Performing joint alignment and image estimation systematically improves motion estimation. Last, we also report the oracle performance of our model with known motions.

Method	Scaling factor / blurring kernel std		
	$\times 2/\sigma = 0.7$	$\times 3/\sigma = 1.2$	$\times 4/\sigma = 1.6$
Single Image SR			
RCAN [50]	29.48	27.30	25.59
IRCNN [49]	29.60	26.89	25.32
USRNet [48]	30.55	27.76	26.18
Video SR			
VSR-DUF[21]	-	31.03	29.24
Ours (no refine)	42.36/0.10	32.63/0.14	**30.00**/0.19
Ours	**43.73**/0.07	**33.10**/0.10	29.87/0.14
Ours (known $\mathbf{p}$)	45.72/0.00	34.47/0.00	31.32/0.00

Table A1: **Results for RGB with synthetic affine motions**, of different methods for different combinations of scale factors and blur kernels. Results are given in term of average PSNR in dBs and geometrical registration error in pixels for our models. "known $\mathbf{p}$" is the oracle performance our model could achieve, if motion estimation was perfect.

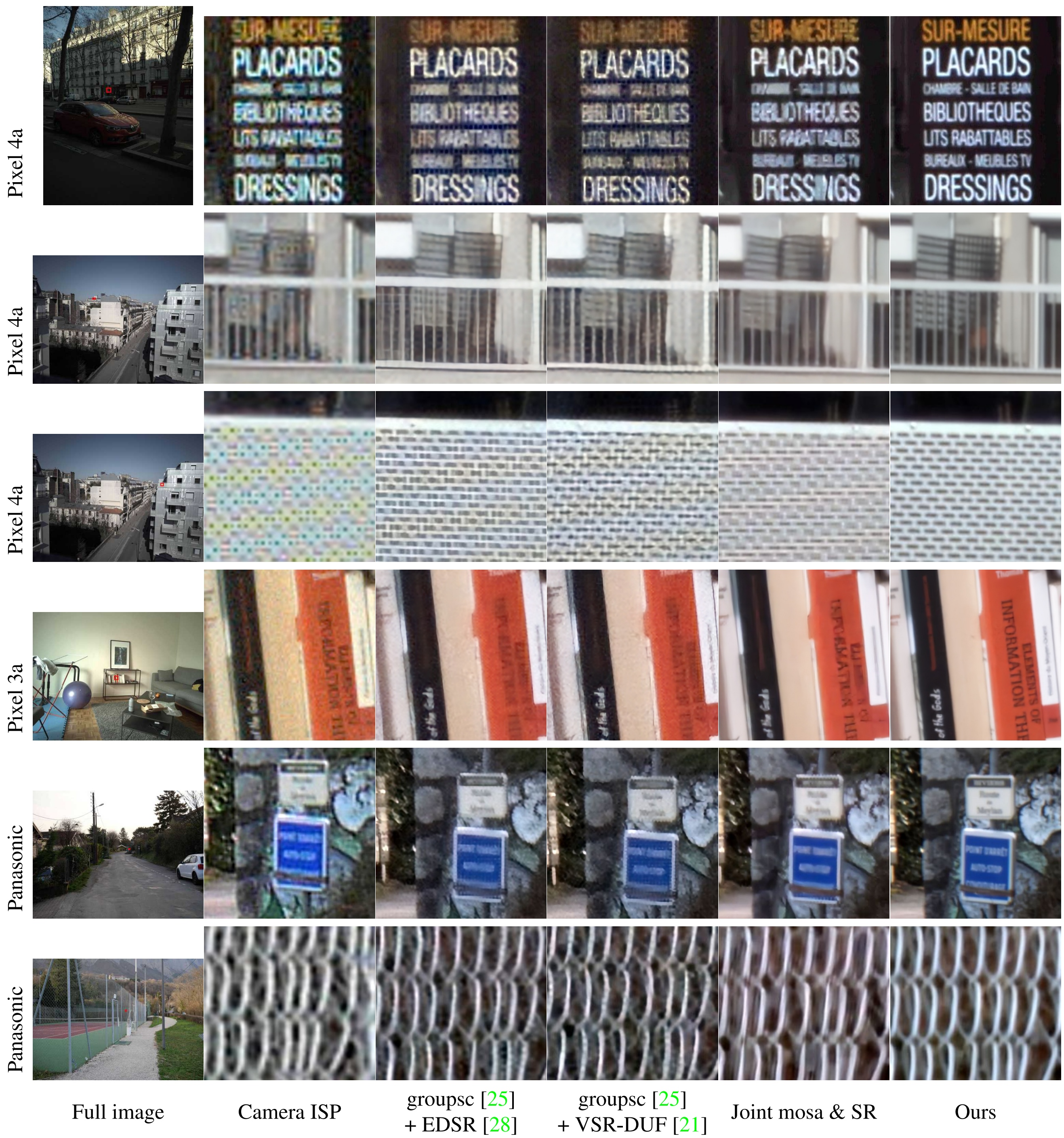
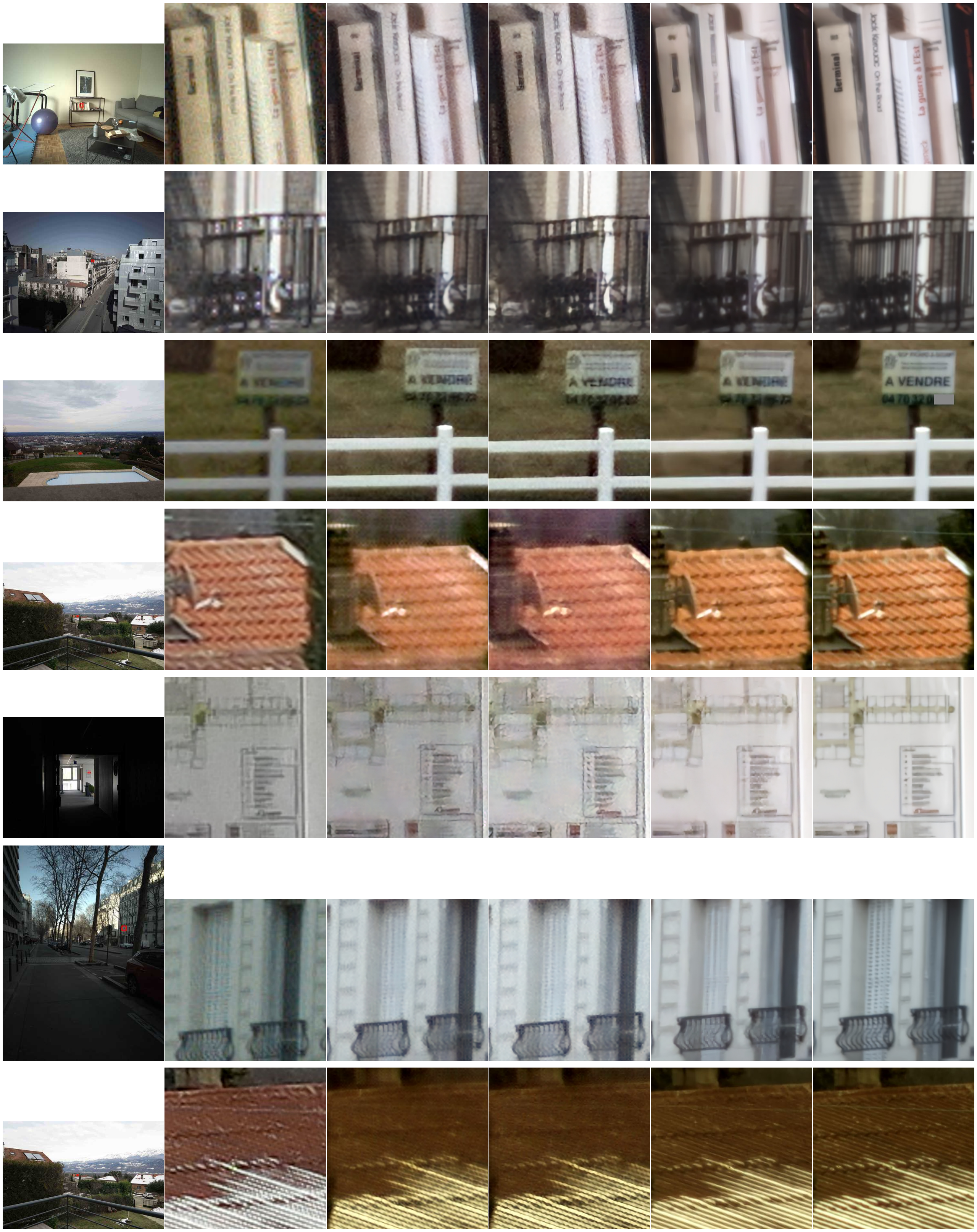


Figure A2: Results from real raw image bursts obtained with various cameras. We provide comparisons with single image and multiframe baselines. Finest restored details can be seen by zooming on computer screen.



Pixel 3a

Pixel 4a

Samsung S7

Canon G7X

Panasonic

Pixel 4a

Canon G7X

Full image | Camera ISP | groupsc [25] + EDSR [28] | groupsc [25] + VSR-DUF [21] | Joint mosa & SR | Ours

Figure A3: Results from real raw image bursts obtained with various cameras. We provide comparisons with single image and multiframe baselines. Finest restored details can be seen by zooming on computer screen.

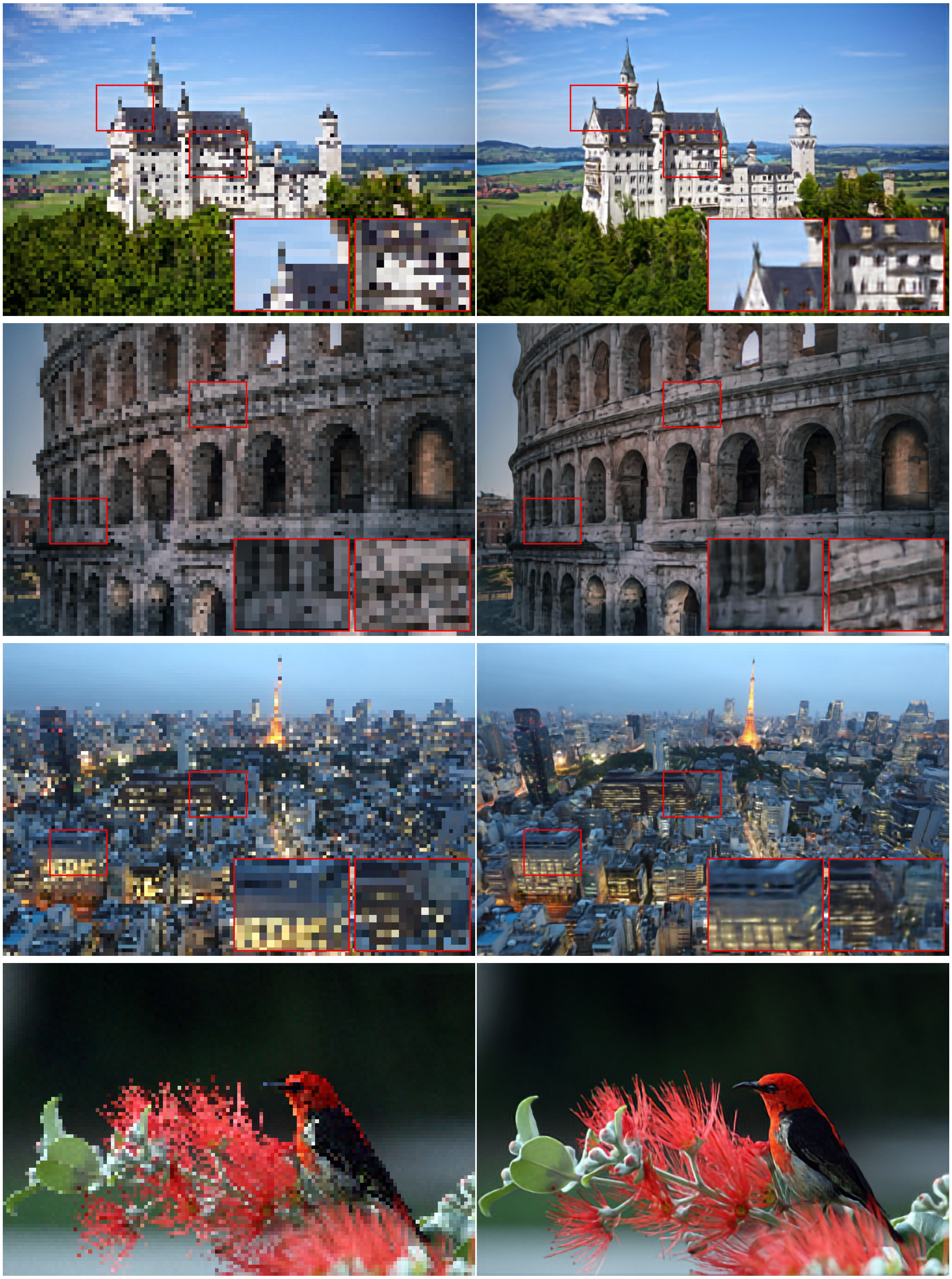


Figure A4: Extreme $\times 16$ upsampling experiment. The right image is obtained by processing a burst of 20 LR images presented on the left obtained with synthetic random affine movements and average pooling downsampling

Figure A5: Image restoration of images taken at night with very low signal to noise ratio by using a Panasonic GX9 camera.

Figure A6: Visual differences caused by merging a different number of frames in the case of low SNR scenes. With a larger number of frames we can observe a quality increase and better denoising.

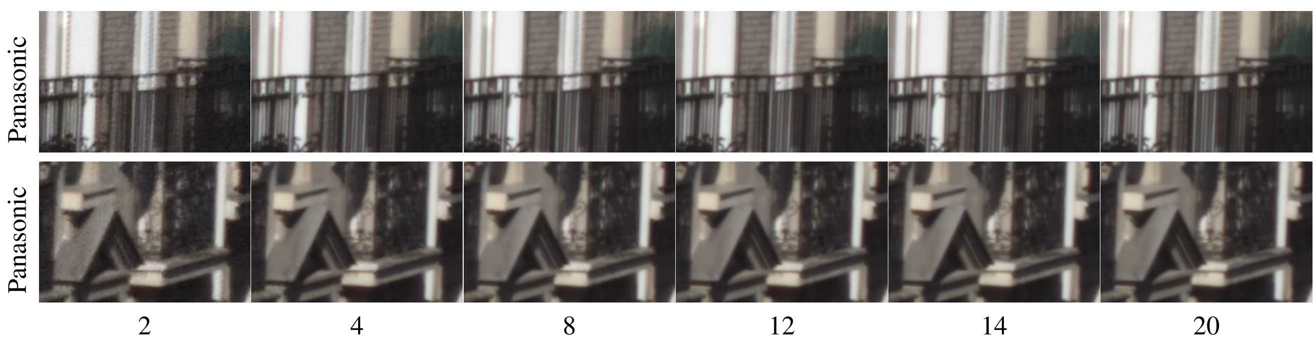


Figure A7: Visual differences caused by merging a different number of frames in the case of high SNR scenes. With a larger number of frames we can observe a quality increase.

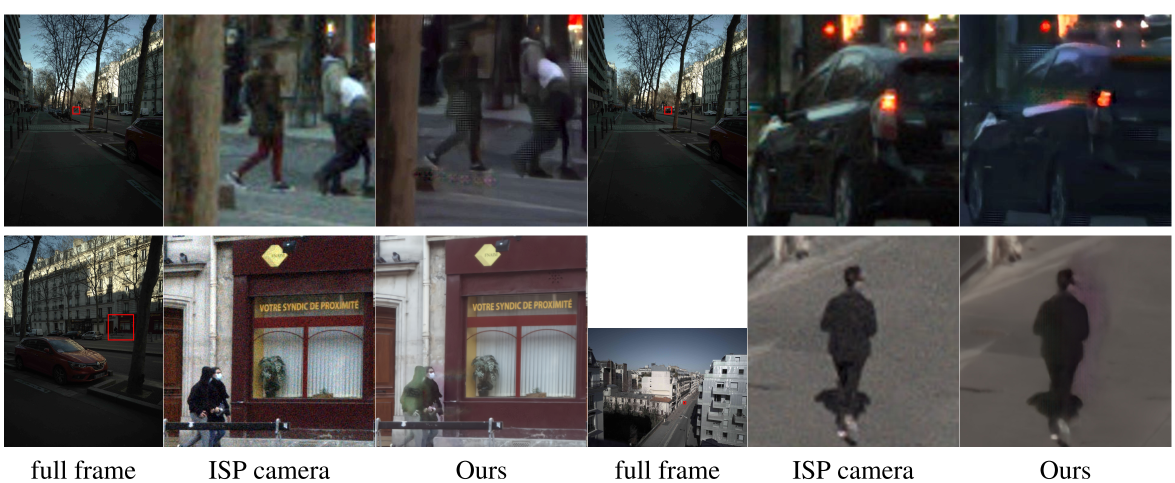


Figure A8: Misalignements artefacts due to moving objects in the scene. Our current implementation does not handle fast moving objects and then generates visual artefacts. Dealing with fast dynamic scenes will be the focus of future work.