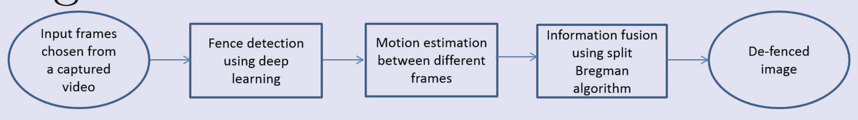
My camera can see through fences: A deep learning approach for image de-fencing

Sankaraganesh Jonna, Krishna K. Nakka, Rajiv R. Sahay
{sankar9.iitkgp, krishkanth.92, sahayiitm}@gmail.com, Indian Institute of Technology Kharagpur, India

Problem

To build an effecient semi-automated system for the detection and removal of fence like occlusions from images/videos using deep learning and sparsity based regularizer.



Model

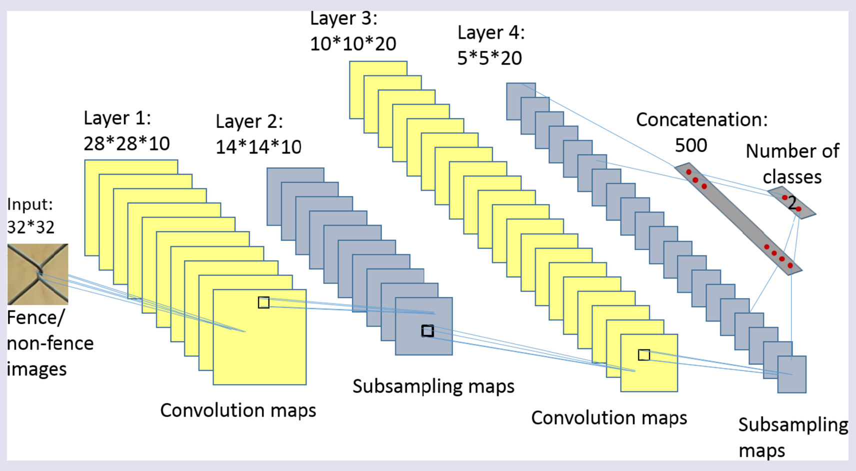
We propose to model the image affected by fences as

$$\mathbf{y}_m = \mathbf{O}_m \mathbf{W}_m \mathbf{x} + \mathbf{n}_m$$

where $\mathbf{y}_m$ represents the m^{th} frame of the video, $\mathbf{O}_m$ is the binary fence mask corresponding to m^{th} frame, $\mathbf{W}_m$ is the warp matrix, $\mathbf{x}$ denotes the de-fenced image and $\mathbf{n}_m$ is the noise assumed as Gaussian.

Detection using CNN

Convolutional neural networks [1] can be effectively trained to recognize objects directly from their images with robustness to scale, shape, angle, noise etc. Hence, we are motivated to investigate the utility of CNNs for the task of fence detection. The overall architecture of the proposed convolutional neural network is described in the following figure.



Dataset for training of the proposed CNN architecture consists of 4000 positive and 8000 negative examples of fence texels. The weights of the proposed CNN are trained by the conventional back-propagation method. The total number of learnable parameters in the proposed CNN architecture is 6282. We have chosen batch size as 50 and constant learning rate α=0.5 throughout all the layers. The trained CNN model is used to classify the fence pixels in the entire test image.

Data fusion using sparsity based regularization

The de-fenced image is the solution of the following constrained optimization problem

$$\arg\min_{\mathbf{x}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x} \|_2^2 + \mu \| \mathbf{d} \|_1 \quad s.t.\ \mathbf{d} = \nabla \mathbf{x}$$

where p is the number of frames chosen from the video and μ is the regularization parameter. We employ the split Bregman iterative framework since the above optimization is a combination of both $l1$ and $l2$ terms hence it is difficult to solve.

$$\arg\min_{\mathbf{x}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x} \|_2^2 + \mu \| \mathbf{d} \|_1 + \frac{\lambda}{2} \| \mathbf{d} - \nabla \mathbf{x} \|_2^2$$

where λ is the shrinkage parameter. The iterates to solve the above equation are as

$$[\mathbf{x}^{k+1}, \mathbf{d}^{k+1}] = \arg\min_{\mathbf{x},\mathbf{d}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x}^k \|_2^2 + \mu \| \mathbf{d}^k \|_1 + \frac{\lambda}{2} \| \mathbf{d}^k - \nabla \mathbf{x}^k + \mathbf{b}^k \|_2^2$$

$$\mathbf{b}^{k+1} = \nabla \mathbf{x}^{k+1} + \mathbf{b}^k - \mathbf{d}^{k+1}$$

We can now split the above problem into two sub-problems as

Sub Problem 1:

$$[\mathbf{x}^{k+1}] = \arg\min_{\mathbf{x}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x}^k \|_2^2 + \frac{\lambda}{2} \| \mathbf{d}^k - \nabla \mathbf{x}^k + \mathbf{b}^k \|_2^2$$

This sub-problem is solved by a gradient descent method.

Sub Problem 2:

$$[\mathbf{d}^{k+1}] = \arg\min_{\mathbf{d}} \mu \| \mathbf{d}^k \|_1 + \frac{\lambda}{2} \| \mathbf{d}^k - \nabla \mathbf{x}^{k+1} + \mathbf{b}^k \|_2^2$$

The above sub-problem can be solved by applying the shrinkage operator as follows

$$\mathbf{d}^{k+1} = shrink(\nabla \mathbf{x}^{k+1} + \mathbf{b}^k, \frac{\lambda}{\mu})$$

$$\mathbf{d}^{k+1} = \frac{\nabla \mathbf{x}^{k+1} + \mathbf{b}^k}{| \nabla \mathbf{x}^{k+1} + \mathbf{b}^k |} * max(| \nabla \mathbf{x}^{k+1} + \mathbf{b}^k | - \frac{\lambda}{\mu}, 0)$$

The update for $\mathbf{b}$ is as $\mathbf{b}^{k+1} = \nabla \mathbf{x}^{k+1} + \mathbf{b}^k - \mathbf{d}^{k+1}$. We tune the parameters μ, λ to obtain the best estimate of the de-fenced image.

Results: Fence Detection

Qualitative comparison of the performance of the method in [2] with the proposed algorithm on our dataset.

Method	Our Database			NRT Database [2]		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Park et al. [2]	0.94	0.26	0.41	0.95	0.46	0.62
Proposed method	0.84	0.96	**0.89**	0.94	0.96	**0.95**

Qualitative comparison of the performance of the method in [2] with the proposed algorithm on PSU NRT dataset.

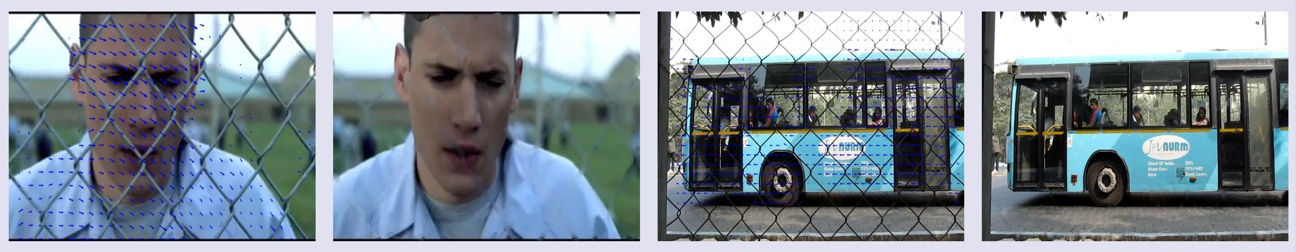
A summary of the quantitative evaluation of the fence detection method of [2] and the proposed algorithm is given in Table.

Results: De-fencing

(a) (b) (c)

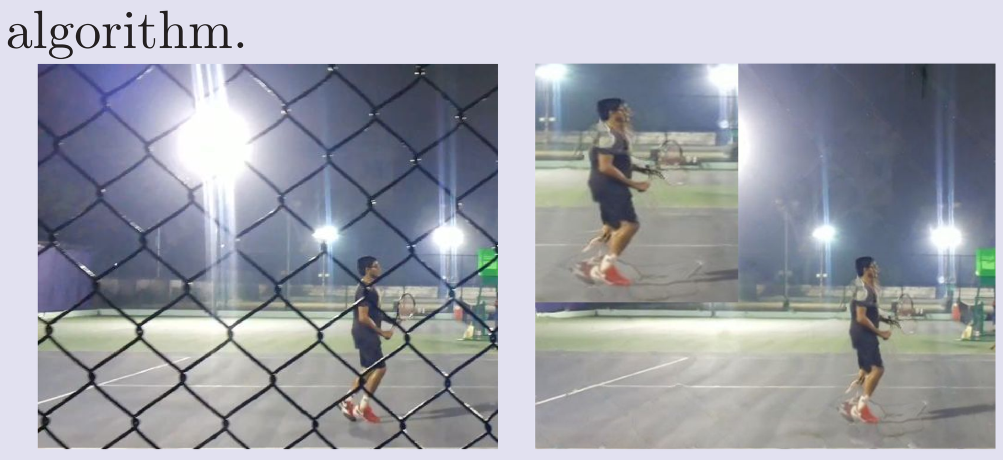
(a) Frame chosen from a video. (b) De-fenced image using proposed algorithm with wrongly estimated motion between frames. (c) De-fenced image obtained using proposed algorithm with accurately estimated motion among frames.

Comparison result with an existing video-based image de-fencing technique in [3].



(a) (b) (c) (d)

(a), (c) are frames chosen from videos. (b), (d) are corresponding de-fenced image obtained using proposed algorithm.



The proposed de-fencing algorithm failed to restore the occluded images if there are any errors in motion estimation between the frames.

References

[1] Y. LeCun and L. Bottou and Y. Bengio and P. Haffner: *Gradient-based learning applied to document recognition*, Proc. IEEE (1998).

[2] M. Park and Brocklehurst, K. and Collins, R. T. and Y. Liu *Deformed Lattice Detection in Real-World Images Using Mean-Shift Belief Propagation*, IEEE Trans. Pattern Anal. Mach. Intell. (2009).

[3] M. Park and K. Brocklehurst and R. T. Collins and Y. Liu *Image De-fencing Revisited*, Asian Conference on Computer vision (2010).

[4] A. Krizhevsky and Sutskever, I. and G. E. Hinton *ImageNet Classification with Deep Convolutional Neural Networks*, Proc. Neural Information and Processing Systems (2012).

Conclusions

- We have proposed an automated CNN based framework for fence detection.
- To validate its accuracy we present a new challenging fenced image dataset consisting of 200 images.
- The split Bregman optimization approach was to obtain the de-fenced image.
- Our results for both synthetic and real-world data show the superiority of the proposed algorithm over the state-of-the-art techniques.