Deep ChArUco: Dark ChArUco Marker Pose Estimation

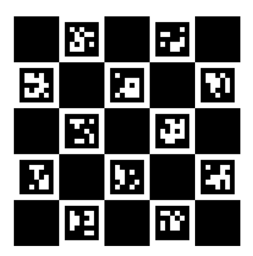
Danying Hu, Daniel DeTone, Vikram Chauhan, Igor Spivak and Tomasz Malisiewicz
Magic Leap, Inc.

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Introduction

In this paper, we present **Deep ChArUco**:

- A deep convolutional neural network system trained to be accurate and robust for ChArUco marker detection under **extreme lighting and motion** and a neural network for subpixel corner refinement
- A novel training dataset collection recipe involving auto-labeling images and synthetic data generation.



Network Architecture

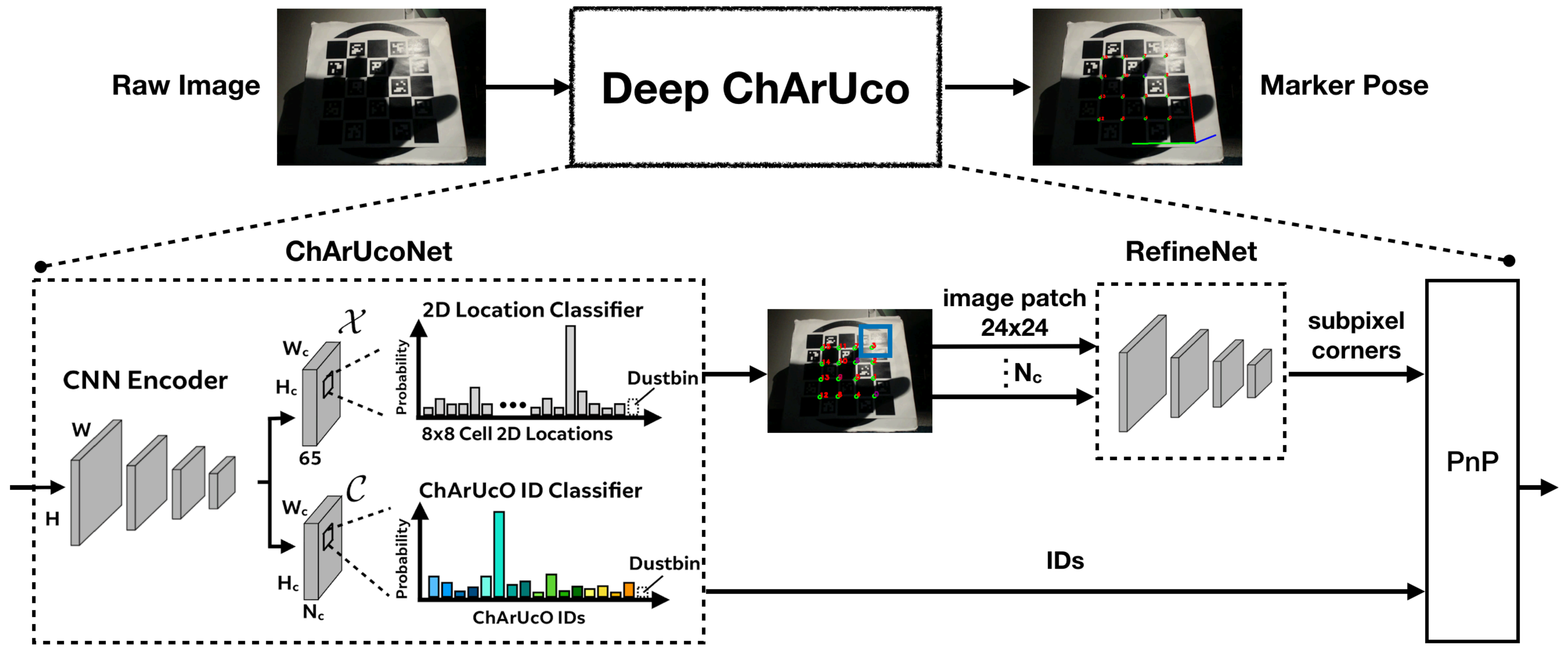


Figure 1: Two-Headed ChArUcoNet and RefineNet. Both **ChArUcoNet** and **RefineNet** are SuperPoint-like [1] networks using VGG-based backbone:

- **ChArUcoNet:** One of the network heads detects 2D locations of ChArUco board's corners and the second head classifies them.
- **RefineNet:** takes a 24×24 image patch and outputs a single subpixel corner location at 8× the resolution of the central 8×8 region.

Training ChArUcoNet

Data generation (see Figure 2)

Data augmentation with synthetics effects:

• blur (gaussian, motion, speckle) • lighting • homographic transform

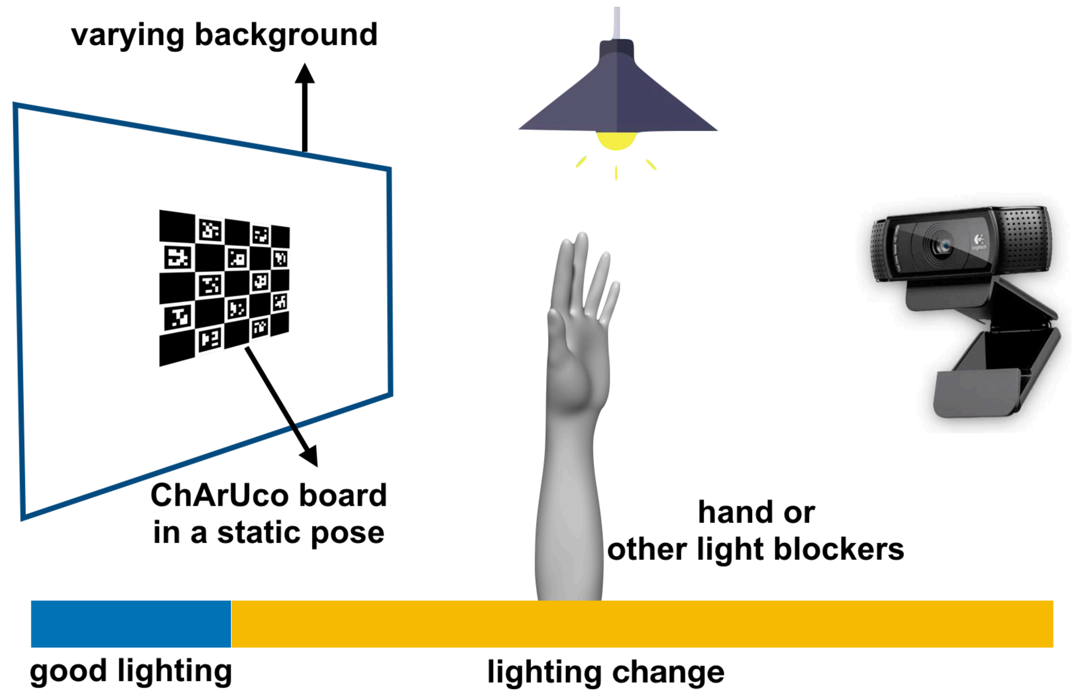


Figure 2: Training data collection

Figure 3: Examples of ChArUco dataset, before and after data augmentation.

Training RefineNet

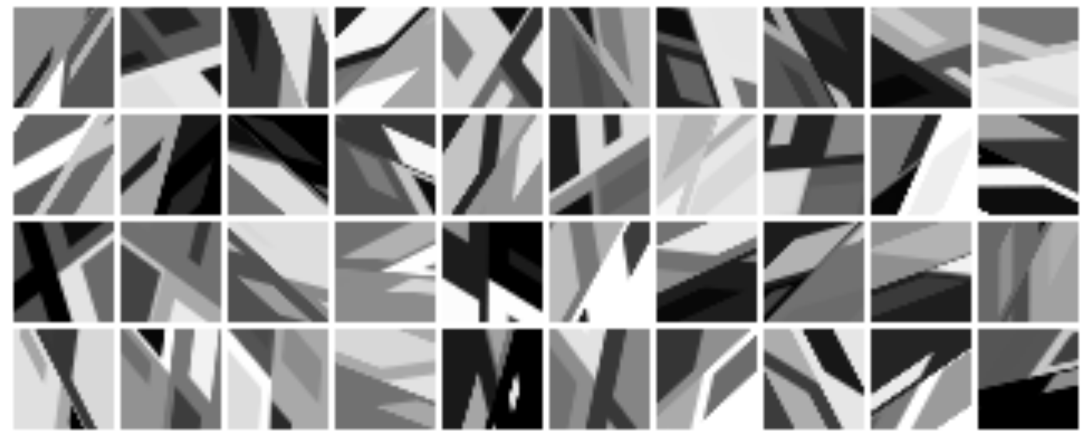


Figure 4: Examples of synthetic training patches. Each image is 24×24 pixels and contains one a ground-truth corner within the central 8×8 pixel region.

Evaluation on synthetic blur/lighting

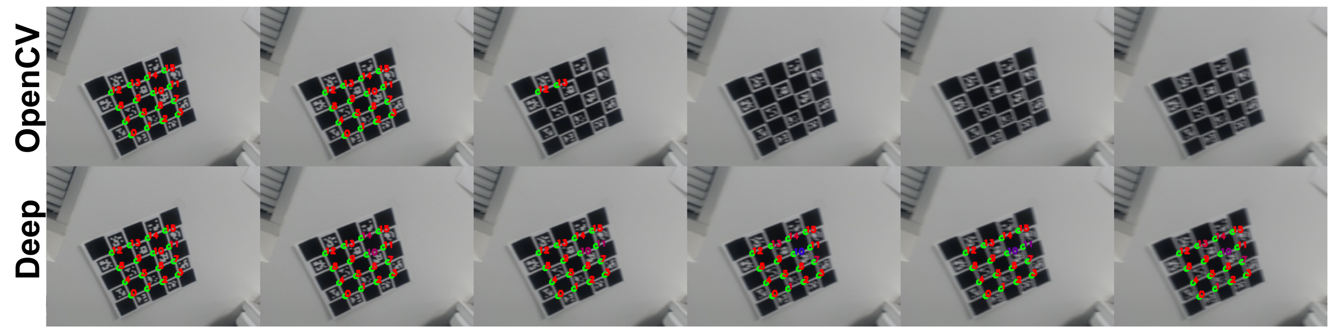
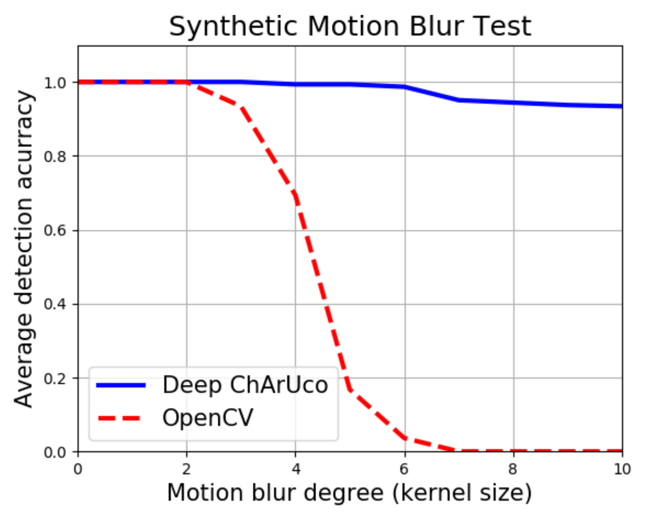


Figure 5: Synthetic motion blur



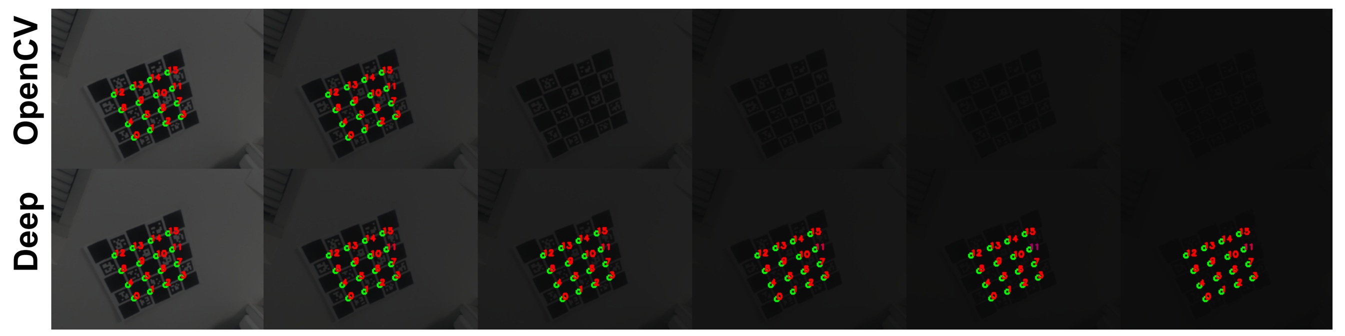
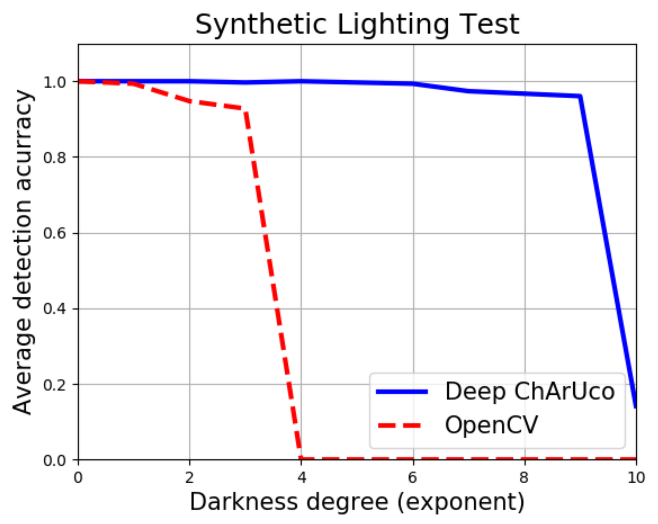


Figure 6: Synthetic lighting



Evaluation on real video sequences

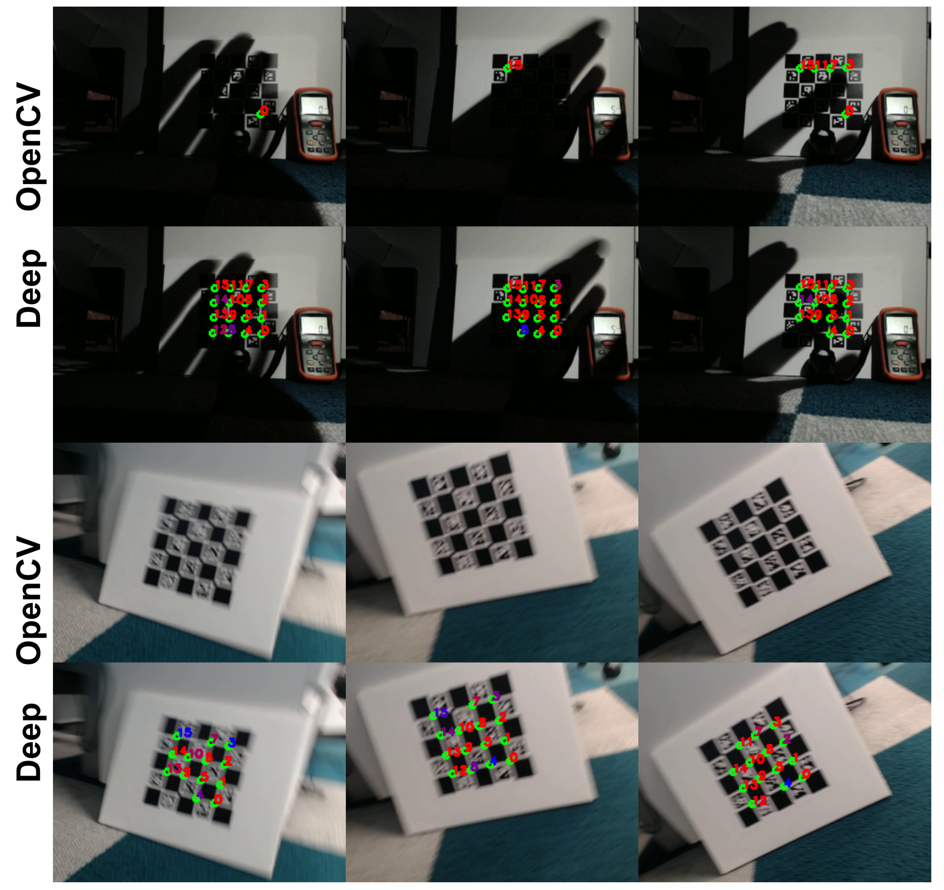
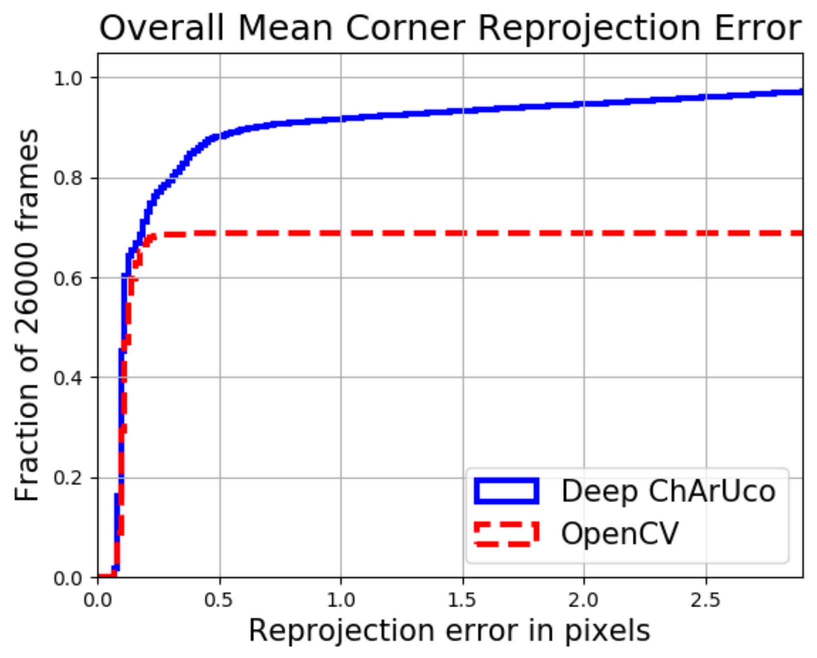


Figure 7: Detector performance comparison under extreme shadows (top) and motion (bottom).



Video	deep acc	cv acc	deep ϵ_{re}	cv ϵ_{re}
0.3lux	**100**	0	**0.427 (0.858)**	nan
0.3lux	**100**	0	**0.388 (0.843)**	nan
1lux	**100**	0	**0.191 (0.893)**	nan
1lux	**100**	0	**0.195 (0.913)**	nan
3lux	100	100	**0.098** (0.674)	0.168
3lux	100	100	**0.097** (0.684)	0.164
5lux	100	100	**0.087** (0.723)	0.137
5lux	100	100	**0.091** (0.722)	0.132
10lux	100	100	**0.098** (0.721)	0.106
10lux	100	100	**0.097** (0.738)	0.105
30lux	100	100	0.100 (0.860)	**0.092**
30lux	100	100	0.100 (0.817)	**0.088**
50lux	100	100	0.103 (0.736)	**0.101**
50lux	100	100	0.102 (0.757)	**0.099**
100lux	100	100	0.121 (0.801)	**0.107**
100lux	100	100	**0.100** (0.775)	0.118
400lux	100	100	**0.086** (0.775)	0.093
400lux	100	100	**0.085** (0.750)	0.093
700lux	100	100	**0.102** (0.602)	0.116
700lux	100	100	**0.107** (0.610)	0.120
shadow 1	**100**	42.0	0.254 (0.612)	0.122
shadow 2	**100**	30.1	0.284 (0.618)	0.130
shadow 3	**100**	36.9	0.285 (0.612)	0.141
motion 1	**74.1**	16.3	1.591 (0.786)	0.154
motion 2	**78.8**	32.1	1.347 (0.788)	0.160
motion 3	**80.3**	31.1	1.347 (0.795)	0.147

Table 1: Individual test video summary of the pose detection rate(percentage of frames with reprojection error less than 3 pixels) as well as the mean reprojection error.

Conclusion

This work demonstrates that deep convolutional neural networks can dramatically improve the detection rate for ChArUco markers in low-light, high-motion scenarios where the traditional ChArUco marker detection tools often fail. We have shown that our Deep ChArUco system, a combination of ChArUcoNet and RefineNet, is significantly more robust to adverse effects such as illumination, blur, and shadows.

[1] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *CVPR Deep Learning for Visual SLAM Workshop*, 2018. [Online]. Available: http://arxiv.org/abs/1712.07629