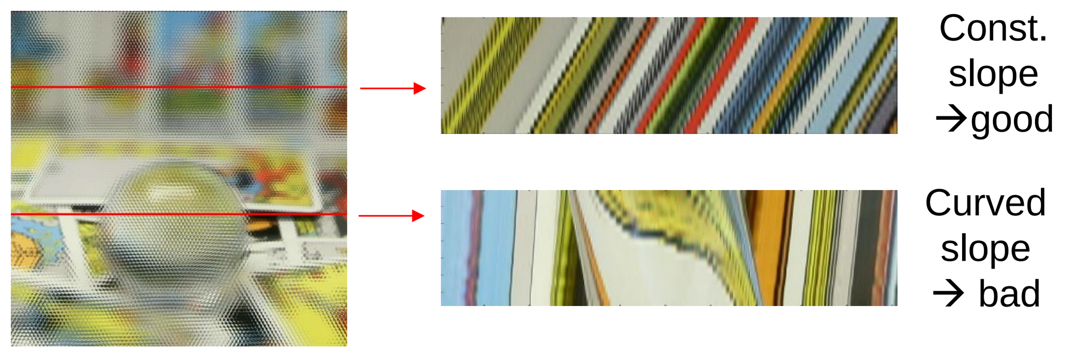
Learning to Detect Light Field Features

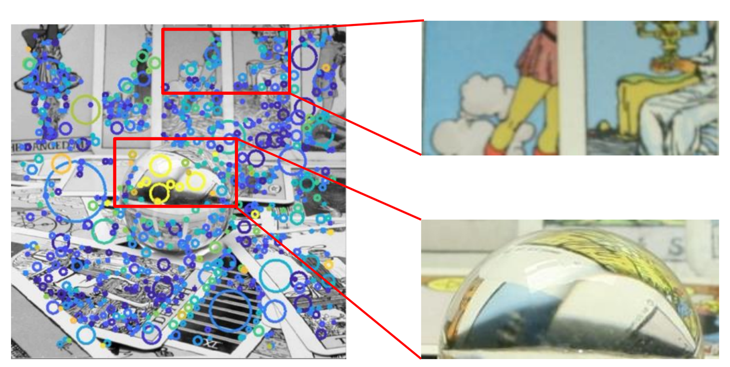
Kelly Guan, Farah Memon, Lars Jebe
Stanford University

EE 368
Winter 2018

Motivation

- Conventional keypoint extraction methods (e.g. SIFT) do not account for additional dimensions in light fields
- Extra information in 4D light fields can **improve quality of detected keypoints**
- A learning approach can significantly **speed up detection**, which is needed for real time applications





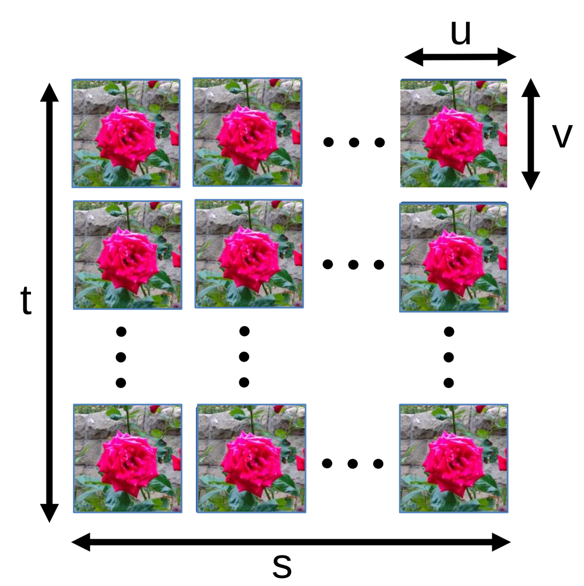
Good features (blue) and bad features (yellow)

Image sources: Stanford Light Field Archive and Donald Danserau

Training Data Acquisition

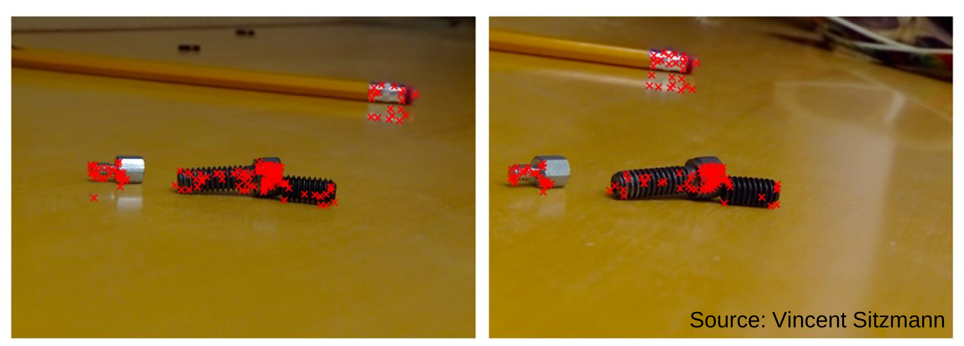
Dataset Specs

Light Field Image Size (u, v, s, t)	541 x 376 x 14 x 14
Rendered Image Size	1404 x 2022
# of Images	4251
# of Views per Scene	4-6
Dataset Size	212 GB
# of Categories	31



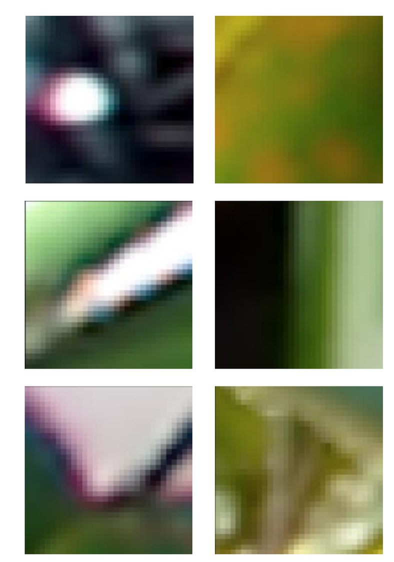
LF Sub-aperture view

- Acquired 8.5 million good patches using SIFT/Harris and bad patches using random selection
- Used COLMAP to filter good features



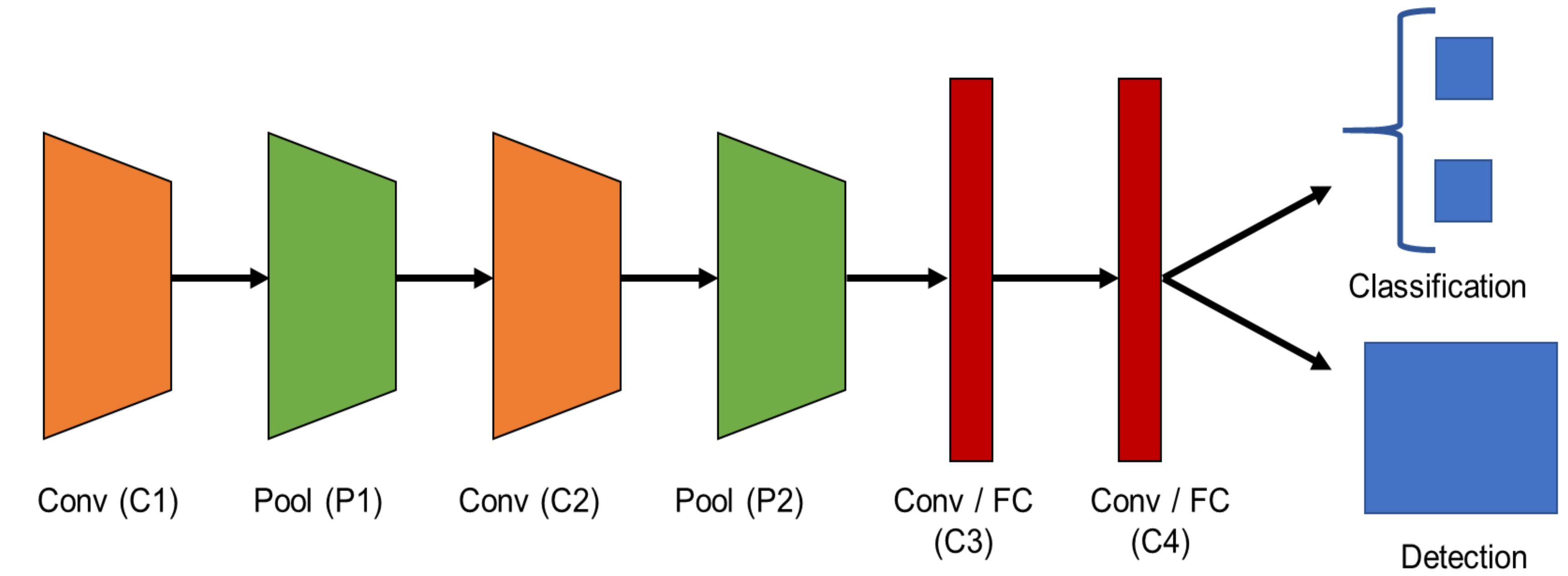
Feature matching using COLMAP

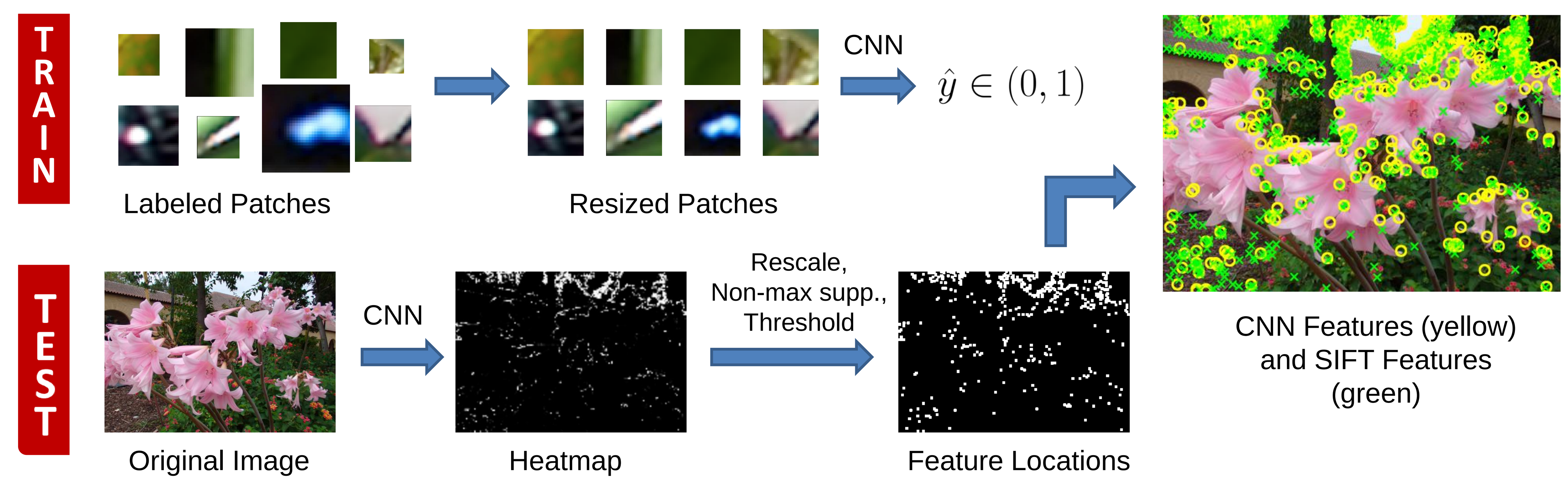
Good Features | Bad Features



Model Architecture and Pipeline

- Binary classification via CNN produces heatmaps for feature detection
- Input: Rendered images (2D CNN) and light field images (3D CNN)

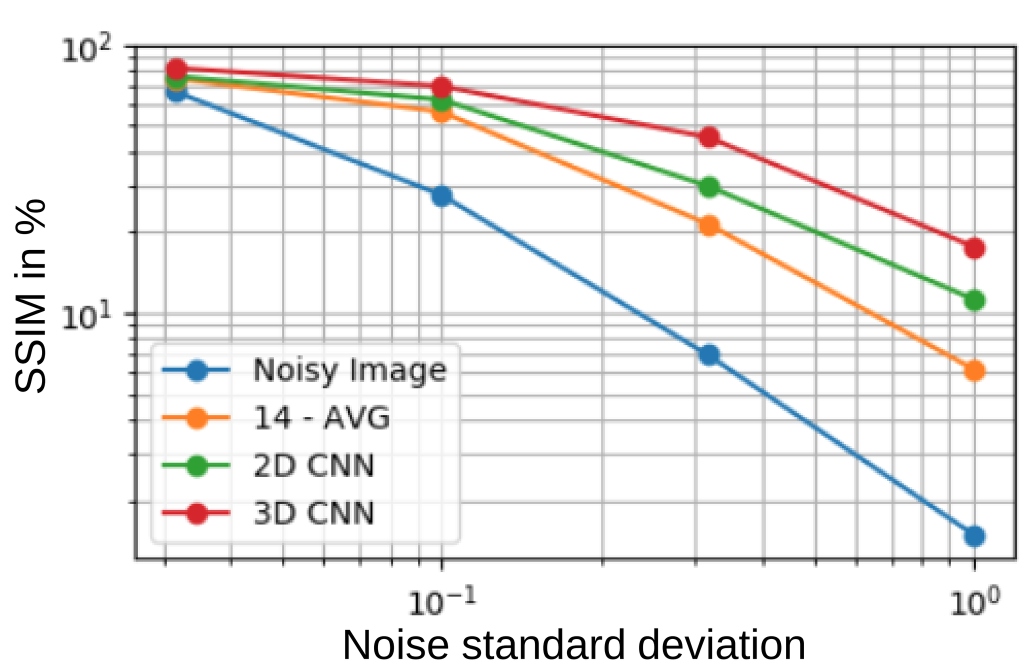




CNN Features (yellow) and SIFT Features (green)

Proof of Concept: Light Field Information

	Original Image	Noisy Image	14 - Average	2D CNN	3D CNN
SSIM:		**8.3 %**	**25.4 %**	**35.1 %**	**57.1 %**



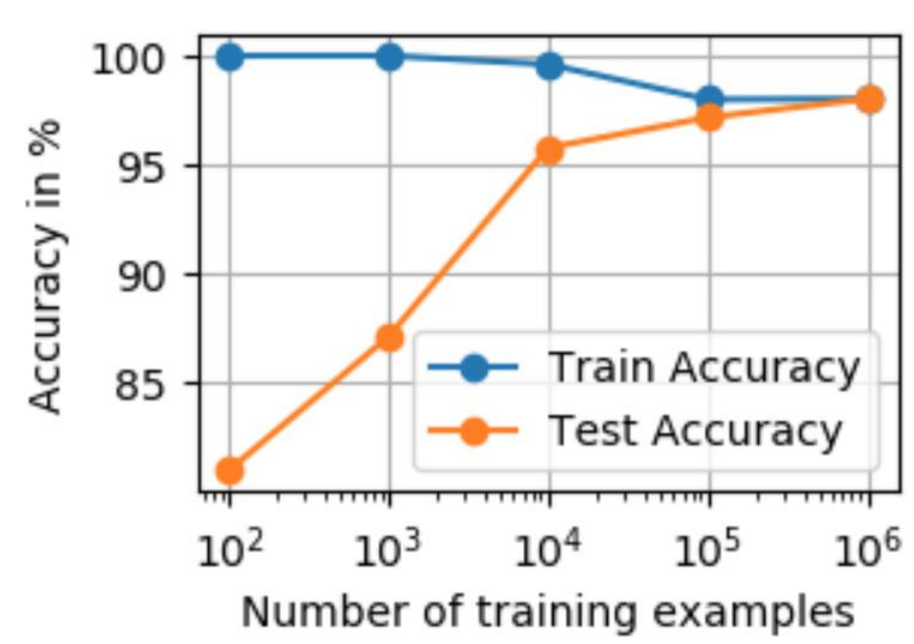
- 3D reconstruction is superior to 2D

→ Additional light field dimension contains useful information

→ Assumption: Useful for feature detection too

Experimental Results

Learning Curve (2D Classification)



- When training on 1 million patches, train and test accuracy converge to ~ 98.6 %

Classification Results

Model	Specifications	Train Acc	Test Acc
2D Classification	Hi-res rendered images	98.7 %	98.6 %
2D Classification LF	2D light field slices	96.0 %	94.7 %
3D Classification LF*	3D light field slices	95.8 %	94.3 %

* Trained on SIFT Features; light field advantage not yet reflected in training data

Detection Results

Model	Specifications	Recall*	Precision*
2D Detection	32 x 32 scale	15.8 %	27.6 %
2D Detection LF	16 x 16 scale	12.2 %	14.7 %

* Compared to SIFT Features of scale 2.67 +/- 1 for 32 x 32 patches and to SIFT features of scale 1.33 +/-1 for 16 x 16 patches

Conclusion and Future Work

Conclusions

- Light field images contain useful information in the additional dimensions
- Classification of patches with CNN works reliably
- Features detection can be learned → good first step
- Light field advantage can be leveraged only with more meaningful training samples

Future Work

- Train model on patches filtered by COLMAP
- Scale input images to detect features at different scales
- Reshape model with recursive layers to resemble scale space within model (Altwaijry et al., 2016)
- Build model that leverages all light field dimensions, e.g. on 3D volume (u,v,depth)

References

David G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. In: International Journal of Computer Vision 60.2 (2004), pp. 91–110.

Xufeng Han et al. “Matchnet: Unifying feature and metric learning for patch-based matching”. In: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. IEEE. 2015, pp. 3279–3286.

Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. “SuperPoint: Self-Supervised Interest Point Detection and Description”. In: arXiv preprint arXiv:1712.07629 (2017).

Hani Altwaijry et al. “Learning to Detect and Match Keypoints with Deep Architectures.” In: BMVC. 2016.

Johannes Lutz Sch¨onberger and Jan-Michael Frahm. “Structure-from-Motion Revisited”. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

Johannes Lutz Sch¨onberger et al. “Comparative Evaluation of Hand-Crafted and Learned Local Features”. In: Conference on Computer Vision and Pattern Recognition (CVPR). 2017.