DEEP FUNDAMENTAL MATRIX ESTIMATION

REN´E RANFTL AND VLADLEN KOLTUN

HAMMAD AYYUBI

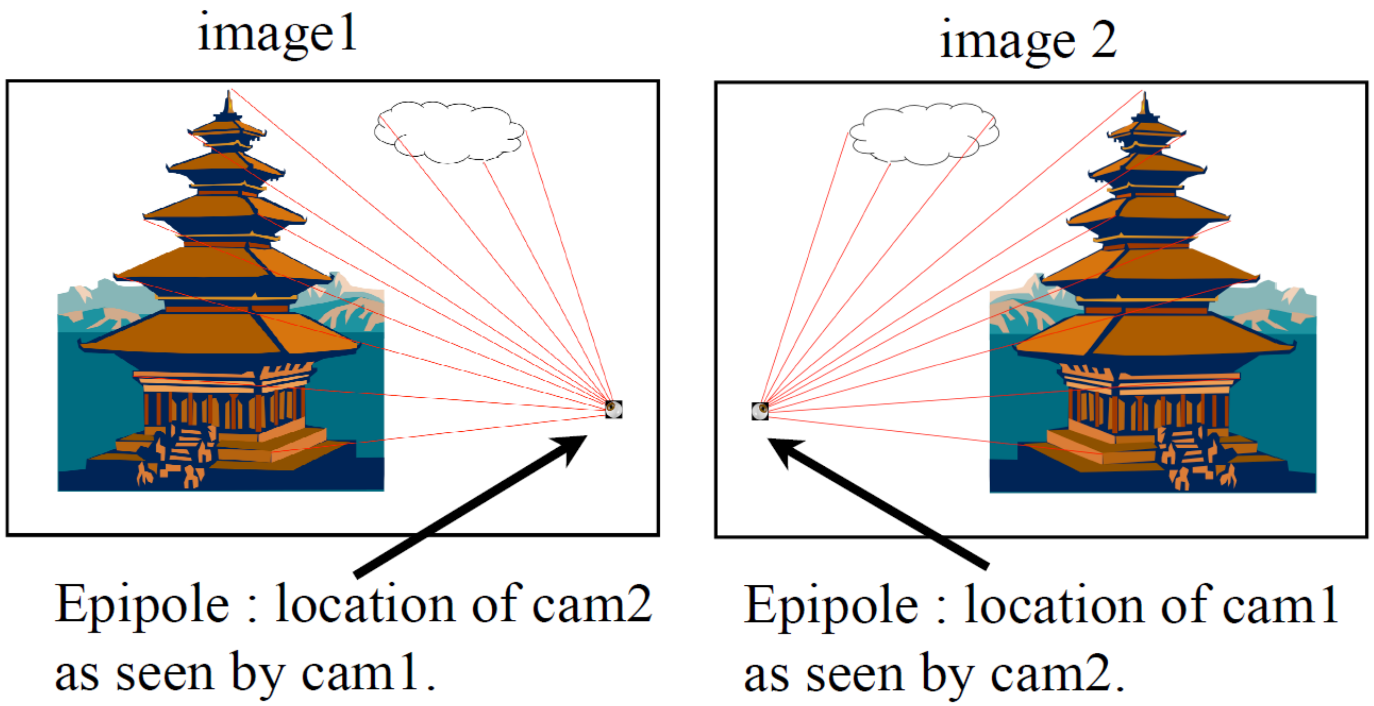
CSE-291D

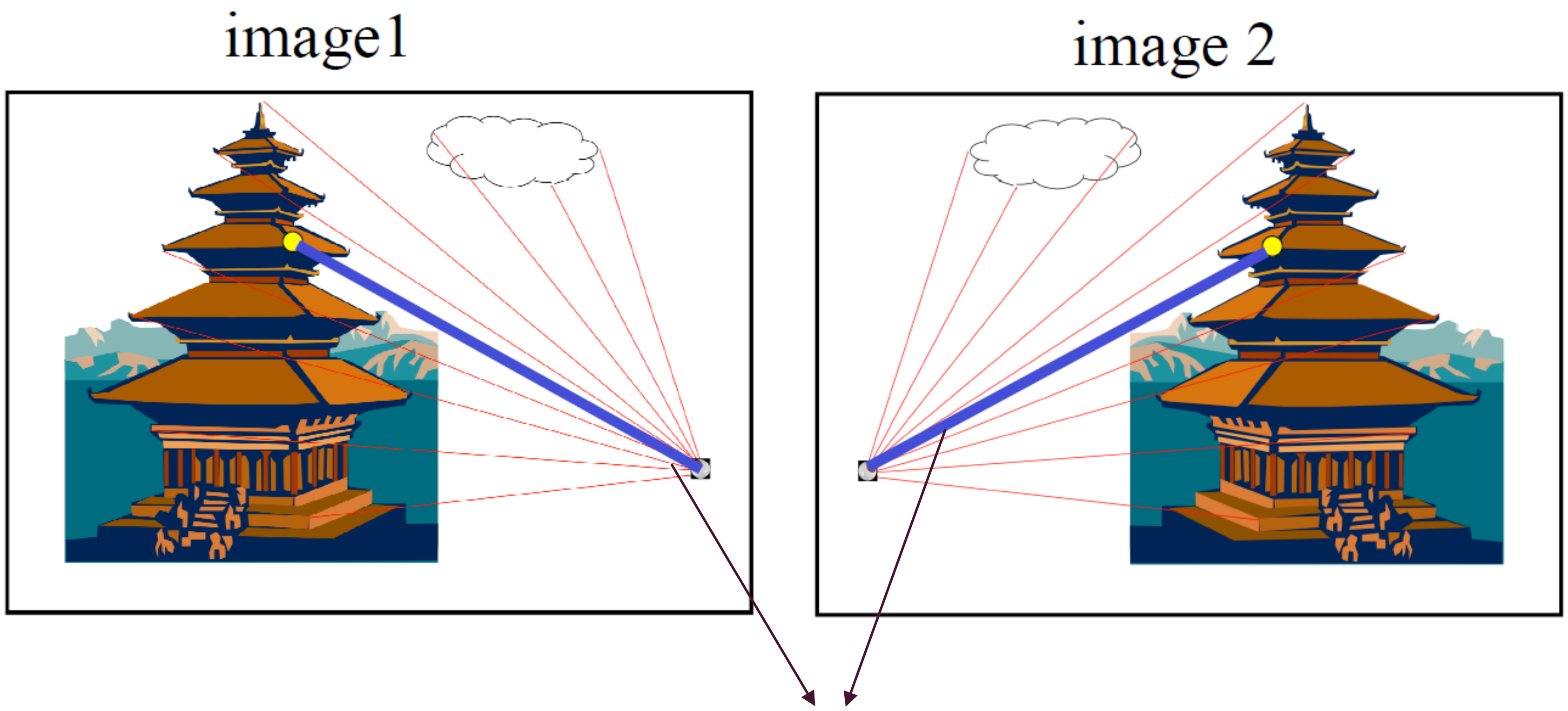
CONTENTS

BACKGROUND AND TOPIC SUMMARIZATION

WHAT IS FUNDAMENTAL MATRIX?

Epipoles and Epipolar Lines





Epiplolar Lines

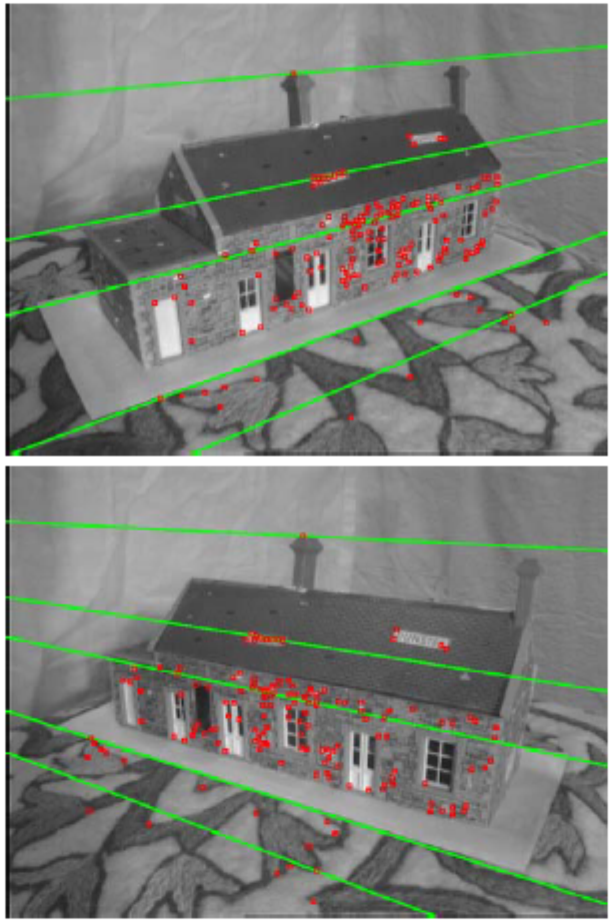
For point x, Fundamental Matrix (F) relates it to the epipolar line l`:

l` = Fx

Key Characteristics of Fundamental Matrix

- It’s a 3x3 matrix
- It has rank 2
- It depends on both extrinsic (R&T) and intrinsic parameters (K) of camera
- Differs from essential matrix in the fact that essential matrix only depends on external parameters

Examples



Fundamental Matrix Estimation

Eight Point Algorithm

$$\begin{bmatrix} x'_i & y'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

$$\begin{bmatrix} x_1x'_1 & x_1y'_1 & x_1 & y_1x'_1 & y_1y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_mx'_m & x_my'_m & x_m & y_mx'_m & y_my'_m & y_m & x'_m & y'_m & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = 0$$

$$A\mathbf{x} = 0$$

$$\min_{\mathbf{x}} ||A\mathbf{x}||^2 \text{ s.t. } ||x||^2 = 1$$

Solution:

- Eigenvector of A with the smallest eigenvalue
- Enforce rank 2 of A by using SVD

Main Challenges in Fundamental Matrix Estimation

- Presence of Outliers/imperfect correspondences
- Different noise distribution in Inliers (Gaussian Assumption in most)
- Huge solution space (Can we narrow it down?)

RELATED WORK

RANSAC

The idea is to find a geometric model that has the most support in terms of inliers. Inliers are defined on the basis of some distance metric from ground truth and a threshold.

Generic steps:

1. A set of points are sampled and a non-robust model is estimated from it.
2. The model is then scored by evaluating a scoring function on all points. The current model is accepted if it's score is better than all previous models.
3. Iterate until a stopping criterion is reached

Shortcoming:

The biggest issue is that as the minimal number of data points required to correctly estimate the model increases, the probability of sampling outlier increases exponentially. For example if half of the correspondences are spurious, the probability of picking up correct ones is only 0.5**8 = 0.39%.

Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography
Tennakoon, R.B., Bab-Hadiashar, A., Cao, Z., Hoseinnezhad, R., Suter, D.: Robust model fitting using higher than minimal subset sampling.

OTHER WORKS

- ✓ Global Optimizations

Yang et al. and Li et al. demonstrate work in which the goal is also achieving consensus set maximaization. But they try to achieve it by optimizing globally. Essentially, pose the problem as Integer Programming Problem.

Shortcoming:

- Since the underlying problem is NP-Hard the process is slow, at times degrading to exhaustive search.

- ✓ Robust base estimators

Robustly estimate base models using a series of weighted least squares problem (Hoseinnezhad et al.).

Shortcomings:

- Careful initialization of weight parameters
- Gaussian Noise assumption in inliers

Li, H.: Consensus set maximization with guaranteed global optimality for robust geometryestimation. In: ICCV (2009)
Yang, J., Li, H., Jia, Y.: Optimal essential matrix estimation via inlier-set maximization. In:ECCV (2014)
Hoseinnezhad, R., Bab-Hadiashar, A.: An M-estimator for high breakdown robust estimation in computer vision. Computer Vision and Image Understanding

TECHNICAL DETAILS

PRELIMNARIES

Many geometric problems can be reduced to a least square fitting problem. Given a set of points $\mathrm{P} \in R^{N\times D}$, model parameters $x \in R^D$, the problem can be modelled as:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \sum_{i=1}^{N} \|(\mathbf{A}(\mathbf{P}))_i \cdot \mathbf{x}\|^2 \\ \text{subject to} \quad & \|\mathbf{x}\| = 1, \end{aligned}$$

$$\mathbf{A} : \mathcal{P} \rightarrow \mathbb{R}^{kN\times d'}$$

$$\begin{bmatrix} x_1x_1' & x_1y_1' & x_1 & y_1x_1' & y_1y_1' & y_1 & x_1' & y_1' & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_mx_m' & x_my_m' & x_m & y_mx_m' & y_my_m' & y_m & x_m' & y_m' & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = 0$$

- An example could be fitting a plane given a set of 3D points.

Challenge: Presence of outliers. The estimated model is off depending on the number of outliers.

Solution: One possible way is to introduce a weighting function which gives low priority to outliers.

$$\mathbf{x}^{j+1} = \underset{\mathbf{x}:\ \|\mathbf{x}\|=1}{\arg\min} \sum_{i=1}^{N} w(\mathbf{p}_i, \mathbf{x}^j) \left\|(\mathbf{A}(\mathbf{P}))_i \cdot \mathbf{x}\right\|^2$$

- The equation doesn't adhere to a closed form in general, thus the solution is arrived at by successively reweighing and estimating the model parameters.
- The problem now is to find the correct weight which can take care of outliers.

DEEP MODEL FITTING

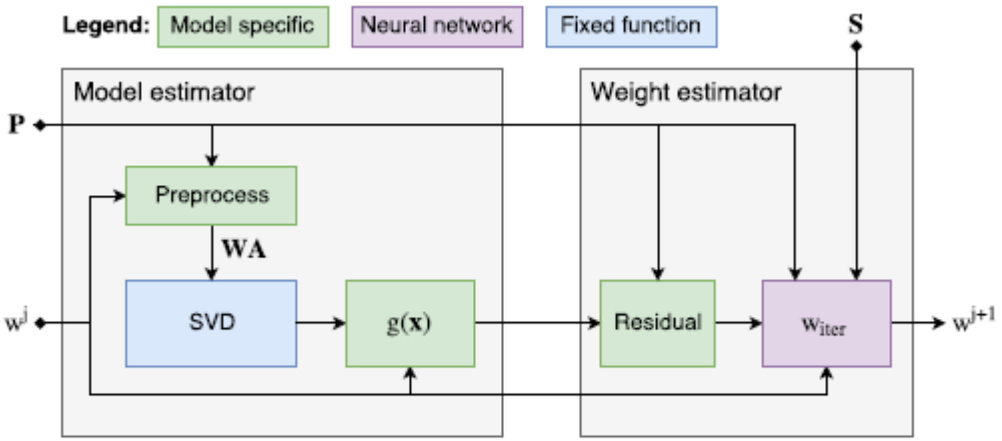
- This work is also based on weighted least squares problem.
- Aims to learn the weighing function/parameter w.

The method is able to perform better in cases where one or more of following is correct:

1. The input data admits regularity in inlier and outlier distribution. Ex: Uniform outlier distribution.
2. Extra information which can be leveraged to further re-weigh the outliers. Ex: Keypoint geometry information.
3. The output space is limited to a narrow region of solution space. Ex: Driving dataset

Technical Details

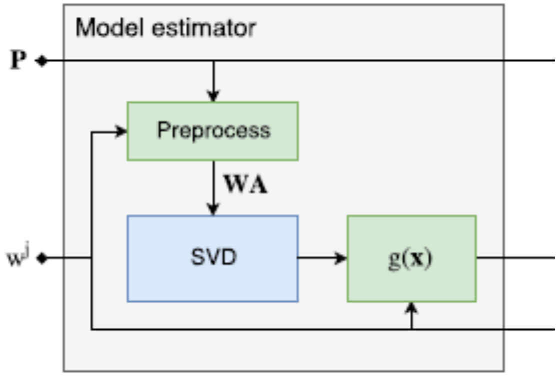
Architecture



Basic building blocks:

1. Model Estimator
2. Weight Estimator

Model Estimator



From our least square weighted equation:

$$\mathbf{x}^{j+1} = \underset{\mathbf{x}:\ \|\mathbf{x}\|=1}{\arg\min} \sum_{i=1}^{N} w(\mathbf{p}_i, \mathbf{x}^j)\, \|(\mathbf{A}(\mathbf{P}))_i \cdot \mathbf{x}\|^2$$

We redefine w according to our deep network – $w : \mathcal{P} \times \mathcal{S} \times \mathbb{R}^{d'} \rightarrow (\mathbb{R}_{>0})^N$
where, $S = R^{N \times s}$ collects side information for each point.

Now, our least square weighted equation becomes:

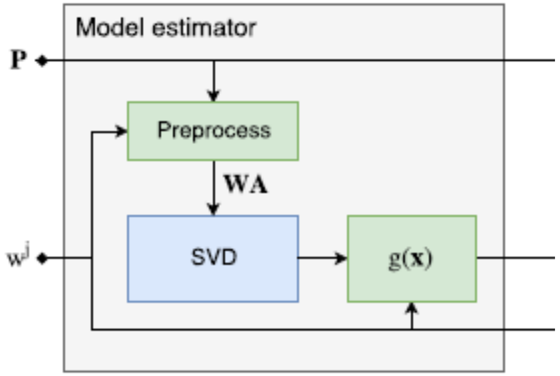
$$\mathbf{x}^{j+1} = \underset{\mathbf{x}:\ \|\mathbf{x}\|=1}{\arg\min} \sum_{i=1}^{N} (w(\mathbf{P}, \mathbf{S}, \mathbf{x}^j; \boldsymbol{\theta}))_i\, \|(\mathbf{A}(\mathbf{P}))_i \cdot \mathbf{x}\|^2$$

$$\mathbf{x}^{j+1} = \underset{\mathbf{x}:\ \|\mathbf{x}\|=1}{\arg\min} \left\|\mathbf{W}^j(\boldsymbol{\theta})\mathbf{A}\mathbf{x}\right\|^2$$

Technical Details

Model Estimator

$$\mathbf{x}^{j+1} = \underset{\mathbf{x}:\ \|\mathbf{x}\|=1}{\arg\min} \left\|\mathbf{W}^j(\boldsymbol{\theta})\mathbf{A}\mathbf{x}\right\|^2$$



Solution:

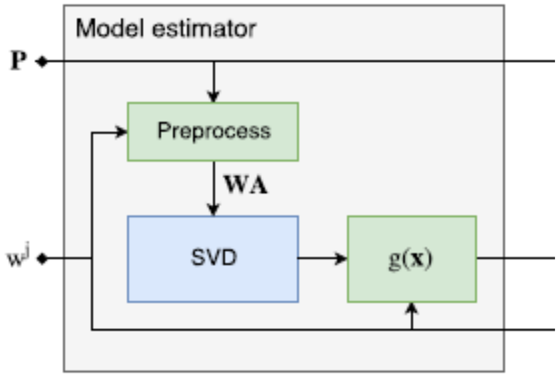
- Eignevector $\grave{v}_d$ of the matrix $W^j(\theta)A$ corresponding to the smallest eigen value.
- One step left – enforcing rank 2 of fundamental matrix. It can be achieved by considering a function $g(x)$ which maps the eigenvector $\grave{v}_d$ to the parameters of the model. (Fundamental Matrix)
- Concisely, we get the model parameters from the matrix $W^j(\theta)A$ by:

$$g(f(\mathbf{W}(\boldsymbol{\theta})\mathbf{A})), \text{ where } f(\mathbf{X}) = \mathbf{v}_{d'}$$

- For training end to end, we need $g(x) and\ f(X)$ to be differentiable.

$$\frac{\partial g}{\partial \mathbf{X}} = \mathbf{U}\left\{2\boldsymbol{\Sigma}\left(\mathbf{K}^\top \circ \left(\mathbf{V}^\top \frac{\partial g}{\partial \mathbf{v}_{d'}}\right)_{sym}\right)\right\}\mathbf{V}^\top, \qquad \mathbf{K}_{ij} = \begin{cases} \frac{1}{\sigma_i^2 - \sigma_j^2}, & if\ i \neq j \\ 0, & otherwise \end{cases}$$

Model Estimator

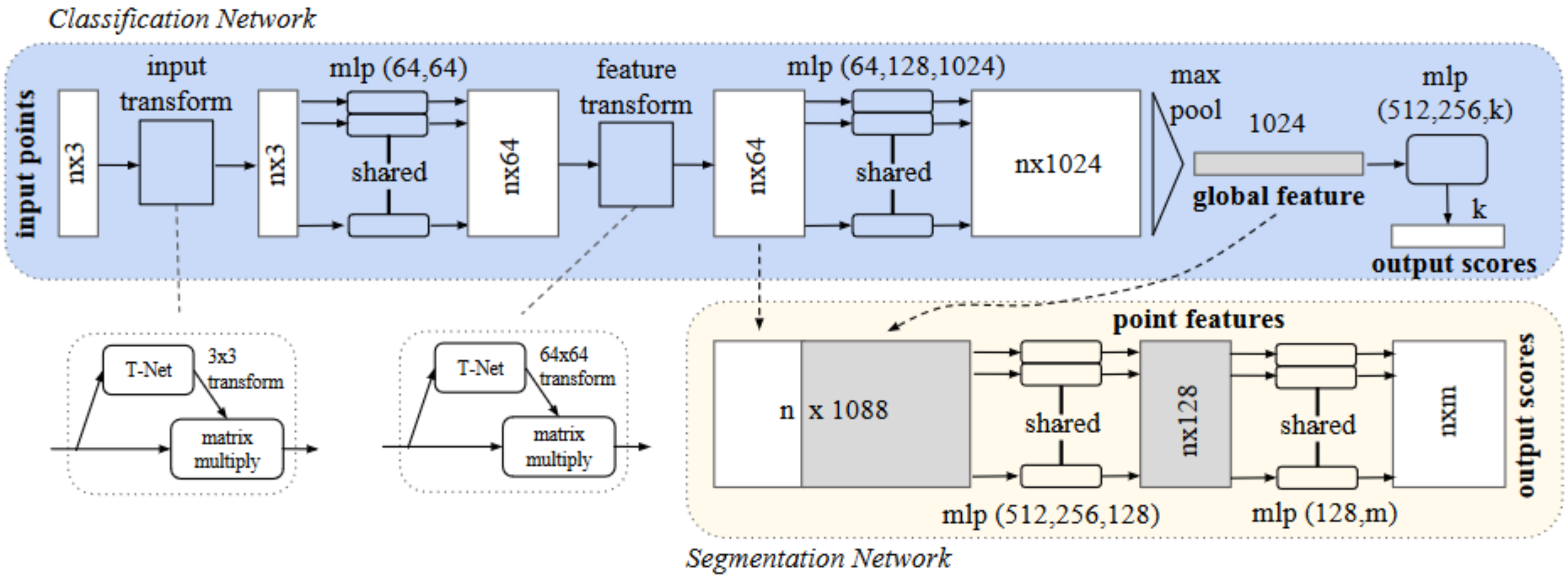


Recap:

- Takes input P and weight w and constructs the matrix $W(\theta)A$
- Applies SVD on $W(\theta)A$ to get the eigenvecor $\grave{v}_d$
- Applies the function $g(x)$ to extract the model parameters (Fundamental Matrix)

Weight Estimator

- The input is a set of points.
- Architecture should be order invariant.
- Based on PointNet

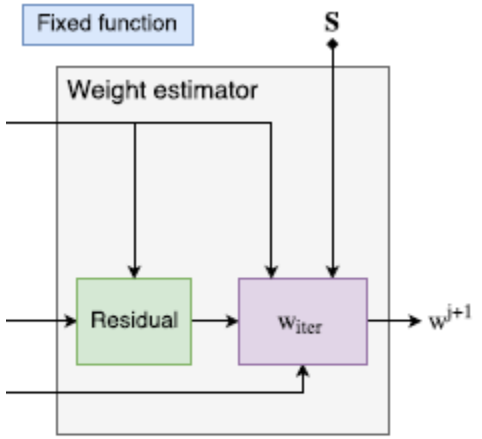


Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep learning on point sets for 3D classification and segmentatio

Weight Estimator

Input:

1. Matrix of points P
2. Side information S for each point
3. Residuals from the current estimate $(\mathbf{r})_i = r(\mathbf{p}_i, g(\mathbf{x}^j))$



Architecture:

1. A linear layer separately for each point.
2. Leaky ReLu as the non-linear function
3. Instance normalization $(I(\mathbf{h}))_i = \frac{\mathbf{h}_i - \mu(\mathbf{h})}{\sqrt{\sigma^2(\mathbf{h}) + \epsilon}},$

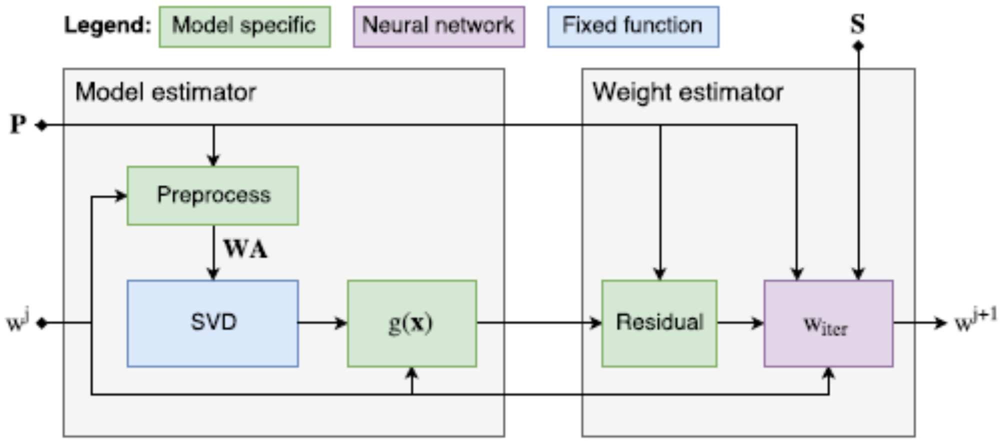
   where, h_i - feature vector for point I

   $\mu(h)$ - mean along dimension N

   $\sigma_2(h)$ - variance along dimension N
4. Final softmax layer gives the weight $w^j(\theta)_i$

Layer	# in	# out	L-ReLU+IN
1	–	64	✓
2	64	128	✓
3	128	1024	✓
4	1024	512	✓
5	512	256	✓
6	256	1	✗

Full Architecture



One Forward Pass:

1. Input weight estimator $w_{init}(\mathbf{P}, \mathbf{S})$
2. Repeated application of Estimation Module (Model Estimator + Weight estimator) – optimal 5.
3. Geometric model estimator on the final weights

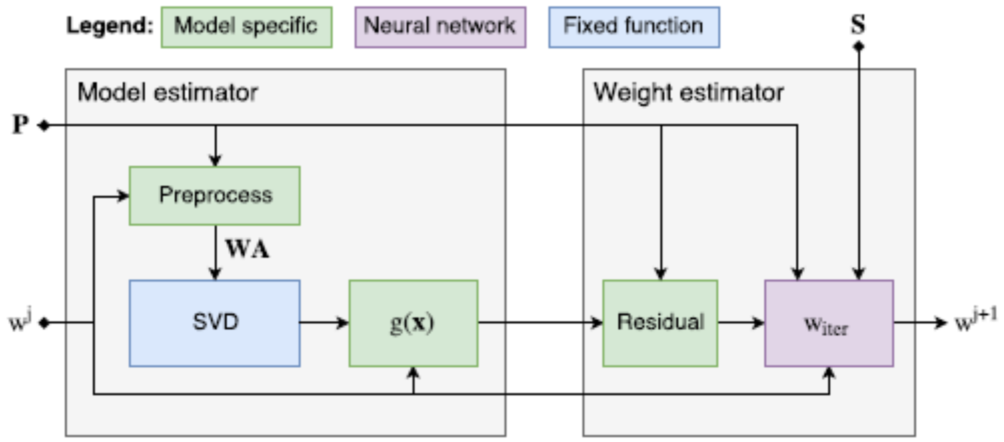
DEEP FUNDAMENTAL MATRIX

Hammad Ayyubi

One Week Later…

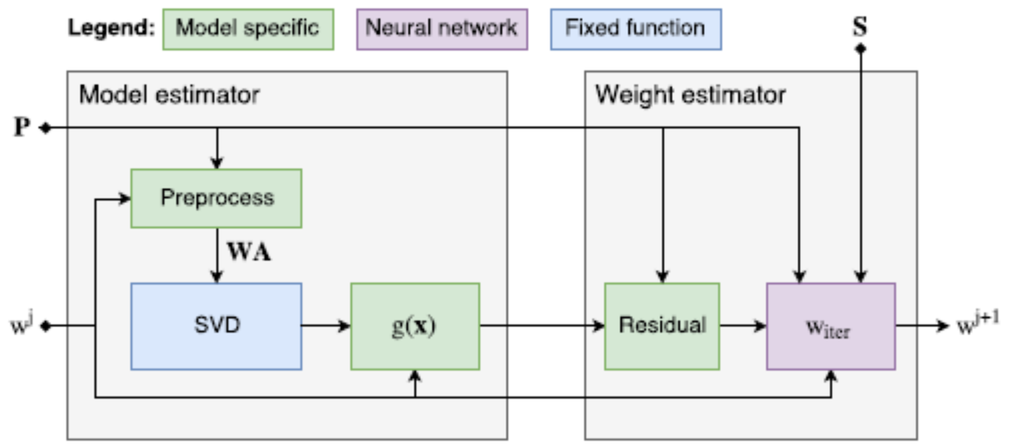
RECAP

1. What is a fundamental Matrix?
2. Means of Estimation – Eight Point Algorithm
3. Main challenges – outliers
4. Different Methods to address it – RANSAC, Global optimization, IRLS
5. The current work builds of IRLS – tries to estimate weight using deep learning
6. Architecture



RECAP

Full Architecture



One Forward Pass:

1. Input weight estimator
2. Repeated application of Estimation Module (Model Estimator + Weight estimator) – optimal 5.
3. Geometric model estimator on the final weights

FUNDAMENTAL MATRIX ESTIMATION

Some module definitions need to be specified:

1. Preprocessing step $A(P)$
2. The model extractor $g(x)$
3. The residual $r(p_i, x)$
4. Training Loss

Fundamental Matrix Estimation

The residual $r(p_i, x)$: $$r(\mathbf{p}_i, \mathbf{F}) = \left|\hat{\mathbf{p}}_i^\top \mathbf{F} \hat{\mathbf{p}}_i'\right| \left(\frac{1}{\|\mathbf{F}^\top \hat{\mathbf{p}}_i\|_2} + \frac{1}{\|\mathbf{F}\hat{\mathbf{p}}_i'\|_2}\right)$$

Symmetric epipolar distance

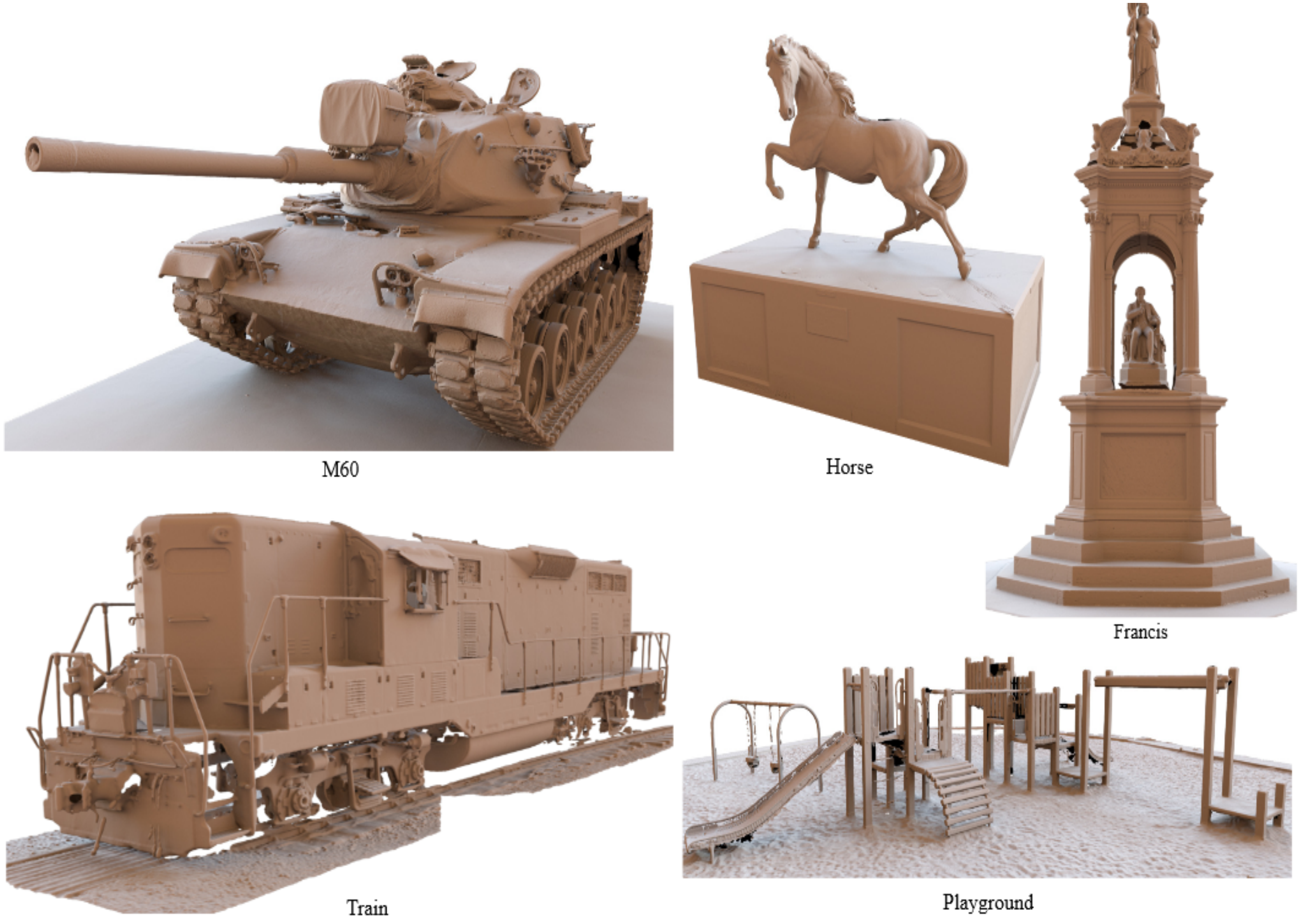
Training Loss: $$\mathcal{L} = \frac{1}{N_{gt}} \sum_{j=0}^{D} \sum_{i=1}^{N_{gt}} \min\left(r(\mathbf{p}_i^{gt}, g(\mathbf{x}^j)), \gamma\right)$$

$\gamma = 0.5$, ensures hard training data doesn't dominate the loss.

EXPERIMENTS

TANKS AND TEMPLES

- Short Image sequences of Train, Horse, M60 etc. taken from hand held camera.
- Use COLMAP SFM (Johannes et. al) to derive fundamental matrices.
- Use SIFT to find potential correspondences; correspondences which are within 1px of epipolar lines are kept as groundtruth.



TANKS AND TEMPLES

- Experiments with varying depths D of estimation module.
- $D = 5^*$ denotes non-usage of side information
- Direct reg. denotes direct regression from points and correspondences to Fundamental matrix estimation (no geometrical information used)

	% Inliers	F-score	Mean	Median	Min	Max	Time [ms]
$D = 1$	42.30	44.80	3.45	1.00	0.08	1912.67	7
$D = 3$	44.91	47.25	1.98	0.82	0.08	566.70	18
$D = 5$	45.02	46.99	2.04	0.83	0.11	285.36	26
$D = 5^*$	44.60	46.42	2.23	0.84	0.10	391.64	26
Direct reg.	4.42	9.14	16.67	11.96	0.83	386.15	3

Inliers: Correspondences with epipolar distance < 1 px
F-score: Positives are those correspondences which have epipolar distance from ground truth epipolar line < 1 px
Mean and Median: Mean and median distances to ground truth epipolar matches.

Experiments

Weight Estimator Module

Layer	# in	# out	L-ReLU+IN
1	–	64	✓
2	64	128	✓
3	128	1024	✓
4	1024	512	✓
5	512	256	✓
6	256	1	✗

Direct Regression

Layer	# in	# out	L-ReLU	Instance norm	Max pooling
1	–	64	✓	✓	✗
2	64	128	✓	✓	✗
3	128	1024	✓	✓	✗
4	1024	512	✓	✓	✗
5	512	256	✓	✓	✓
6	256	128	✓	✗	✗
7	128	64	✓	✗	✗
8	64	9	✓	✗	✗

Comparison with other methods:

- Two experiments: with and without ratio test. Ratio test is used in SIFT to increase % of inliers.

	Tanks and Temples – with ratio test				*Tanks and Temples – without ratio test*			
	% Inliers	F-score	Mean	Median	% Inliers	F-score	Mean	Median
RANSAC	42.61	42.99	**1.83**	1.09	2.98	10.99	122.14	79.28
LMEDS	42.96	40.57	2.41	1.14	1.57	4.78	120.63	108.72
MLESAC	41.89	42.39	2.04	1.08	2.13	8.28	131.11	93.04
USAC	42.76	43.55	3.72	1.24	4.45	23.55	46.32	8.52
Ours	**45.02**	**46.99**	2.04	**0.83**	**5.62**	**26.92**	**36.81**	**7.82**

KITTI

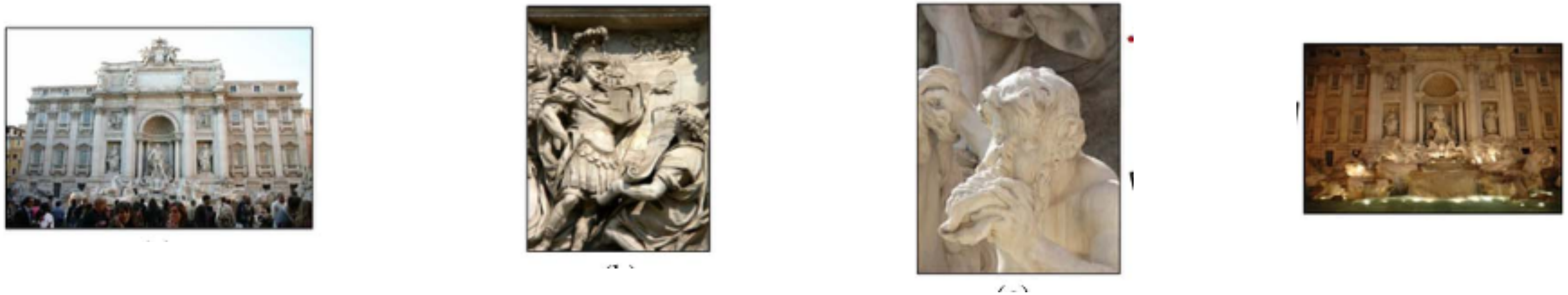
- 22 sequences of driving dataset
- 11 of them has groundtruth odometry.

KITTI

	@ *0.1px*		@ *1px*			
	% Inliers	F-score	% Inliers	F-score	Mean	Median
RANSAC	21.85	13.84	84.96	75.65	0.35	0.32
LMEDS	20.01	13.34	84.23	75.44	0.37	0.35
MLESAC	18.60	12.54	84.48	75.15	0.39	0.36
USAC	21.43	13.90	85.13	75.70	0.35	0.32
Ours tr. on T&T	21.00	13.31	84.81	75.08	0.39	0.33
Ours tr. on KITTI	**24.61**	**14.65**	**85.87**	**75.77**	**0.32**	**0.29**

Even the model trained on T&T achieves comparable performance with the state of the art networks. This demonstrates the generalization ability of the proposed method.

COMMUNITY PHOTO COLLECTION



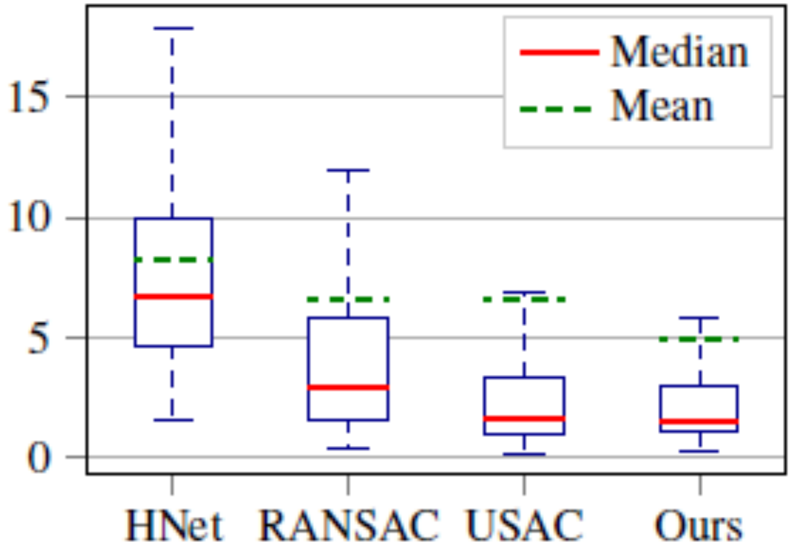
- Dataset contains images taken from variety of cameras and from varying positions.
- The data has no apparent regularity

COMMUNITY PHOTO COLLECTION

	@ 0.1px		*@ 1px*			
	% Inliers	F-score	% Inliers	F-score	Mean	Median
RANSAC	49.55	40.80	67.52	59.12	2.29	1.21
LMEDS	**51.74**	41.87	67.85	59.38	2.50	1.16
MLESAC	48.07	40.01	67.40	58.64	**1.45**	1.17
USAC	51.21	41.87	66.65	58.93	2.94	1.22
Ours	51.41	**43.28**	**68.31**	**60.67**	1.51	**1.02**

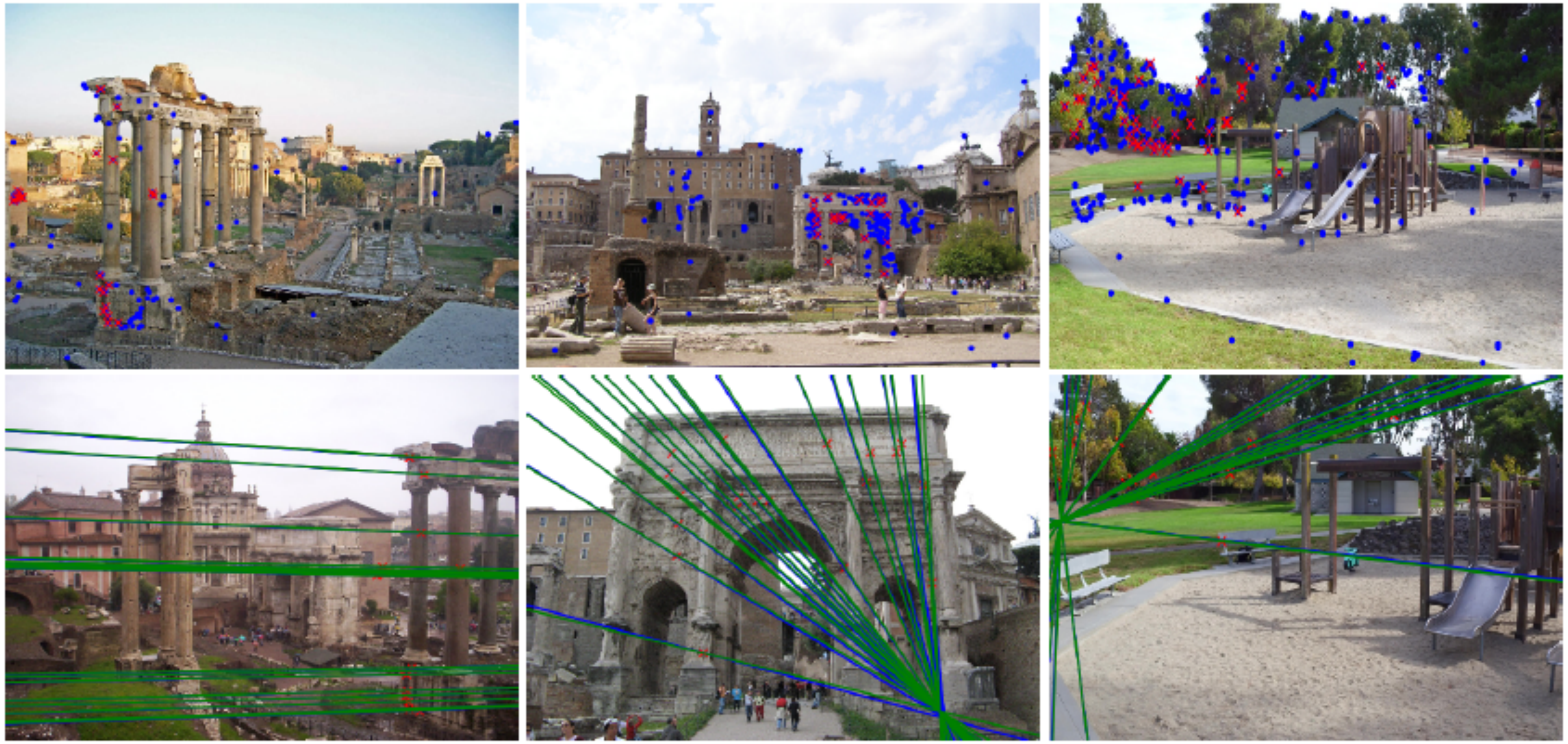
- The data has no apparent regularity
- Good performance indicates good generalization Ability

HOMOGRAPHY ON MS-COCO



- Good performance shows the generic nature of proposed algorithm – can be adapted for any geometrical modelling task based on Direct Linear Transorm

SOME EXAMPLES



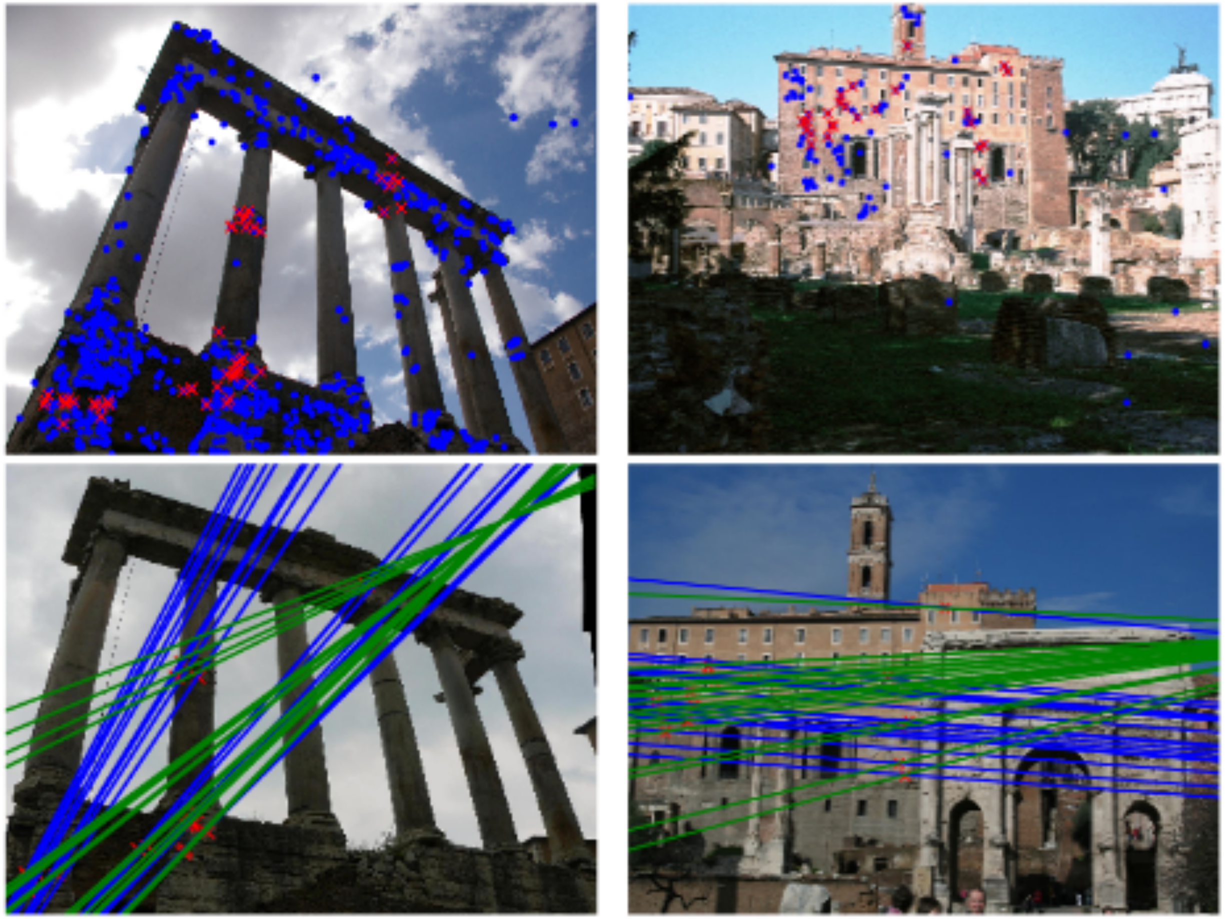
- Blue Dot: Outlier
- Red Dot: Inlier
- Green Line: Predicted
- Blue line: Groundtruth

FAILURE EXAMPLES

What might be the reason for wrong prediction of epipolar lines?

- ✓ Failure to correctly estimate inliers/epipolor lines. Closely look at inlier at bottom.

FAILURE EXAMPLES



- ✓ Left example: Presence of noise
- ✓ Right example: Indistinct features

CRITIQUE

STRENGTHS

1. Demonstrate the advantages of integration of geometrical information into the deep learning pipeline. Also, pure data driven pipelines require huge labelled data which may not be available for a number of 3D reconstruction tasks.
2. The proposed method can learn models specific to data at hand.
3. It is robust to inlier noise distribution.
4. Good generalization ability to multiple datasets.
5. Generic nature of the proposed method encourages application in other 3D tasks which are based on Direct Linear Transform

WEAKNESS

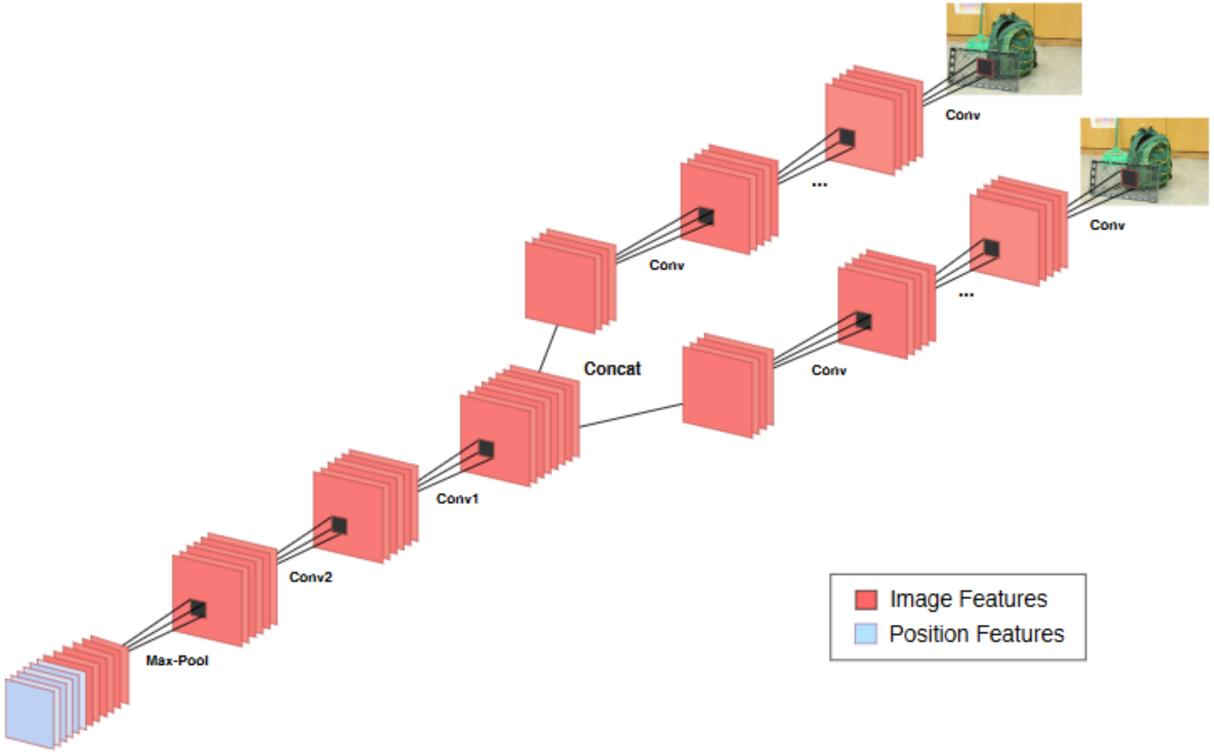
1. The proposed method is still dependent on the accuracy/correctness of correspondence estimator.
2. Accuracy of Fundamental Matrix estimation increases with the number of estimation modules D, but so does the time of evaluation.

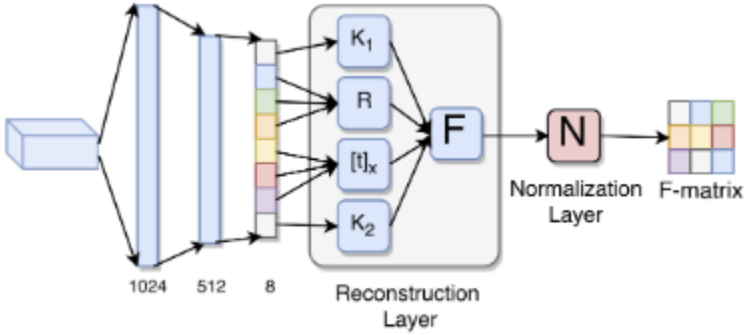
EXTENSIONS AND FOLLOW-UPS

EXTENSIONS AND FOLLOW-UPS

1. Estimating Fundamental Matrix without requiring correspondence estimation from a third party.

Deep Fundamental Matrix Estimation without Correspondences : Omid Poursaeed, Guandao Yang, Aditya Prakash, Qiuren Fang, Hanqing Jiang, Bharath Hariharan, Serge Belongie : ECCV 18





THANK YOU!!
QUESTIONS?

Fundamental Matrix Estimation

Preprocessing step $A(P)$:

$$(\mathbf{A}(\mathbf{P}))_i = \text{vec}\left(\mathbf{T}\hat{\mathbf{p}}_i(\mathbf{T}'\hat{\mathbf{p}}_i')^\top\right)$$

where, $\hat{\mathbf{p}}_i = ((\mathbf{p}_i)_1, (\mathbf{p}_i)_2, 1)^\top$ and $\hat{\mathbf{p}}_i' = ((\mathbf{p}_i)_3, (\mathbf{p}_i)_4, 1)^\top$ are homogenous coordinates of correspondences

$T \text{ and } T'$ are noramalization matrices

The model extractor $g(x)$:

$$g(\mathbf{x}) = \underset{\mathbf{F}:\ \det(\mathbf{F})=0}{\arg\min} \left\|\mathbf{F} - \mathbf{T}^\top(\mathbf{x})_{3\times3}\mathbf{T}'\right\|_F$$

Essentially enforces rank=2 for fundamental matrix F.