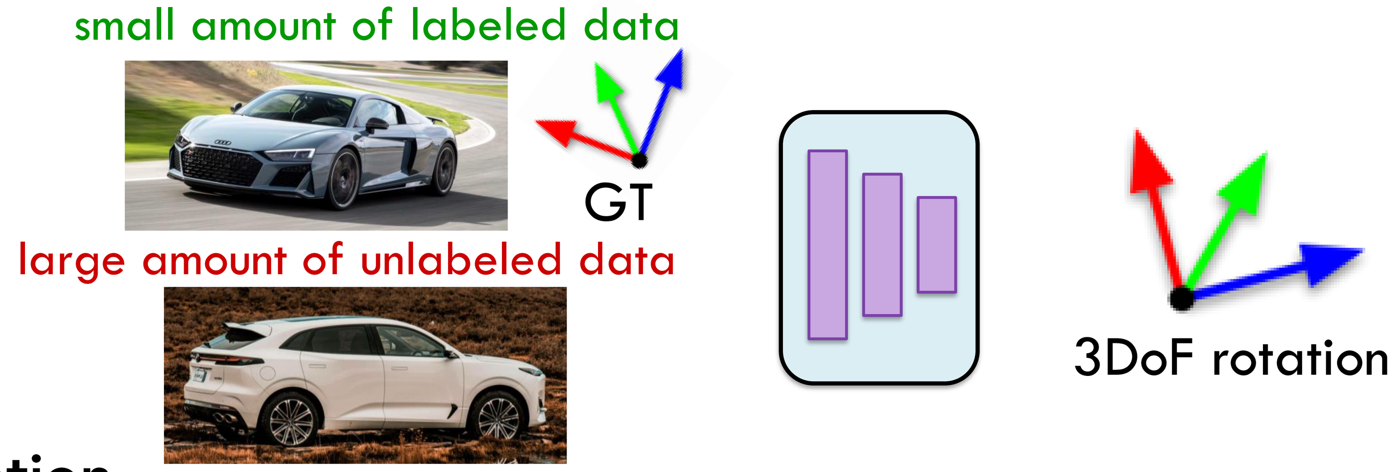
FisherMatch: Semi-Supervised Rotation Regression via Entropy-based Filtering

Yingda Yin, Yingcheng Cai, He Wang†, Baoquan Chen†
Peking University
†: corresponding author

Introduction

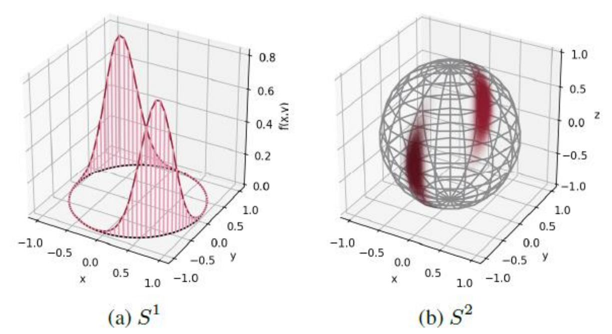
- **Task**: Given **a small number of labeled data** and **a large collection of unlabeled data**, our algorithm learns to regress 3D rotation in a **semi-supervised** manner.

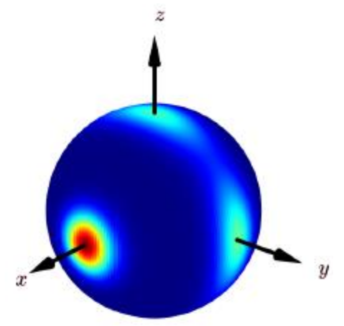


- **Motivation**
  - Very high cost for rotation annotation for 2D images
  - Huge amounts of unannotated data available
- **Our work**
  - The first general semi-supervised rotation regression framework
  - Demonstrates superior efficacy that can learn from few labeled data

Probabilistic Modeling of Rotation

- **Bingham distribution** (for quaternions)
- **Matrix Fisher distribution** (for rotation matrices)

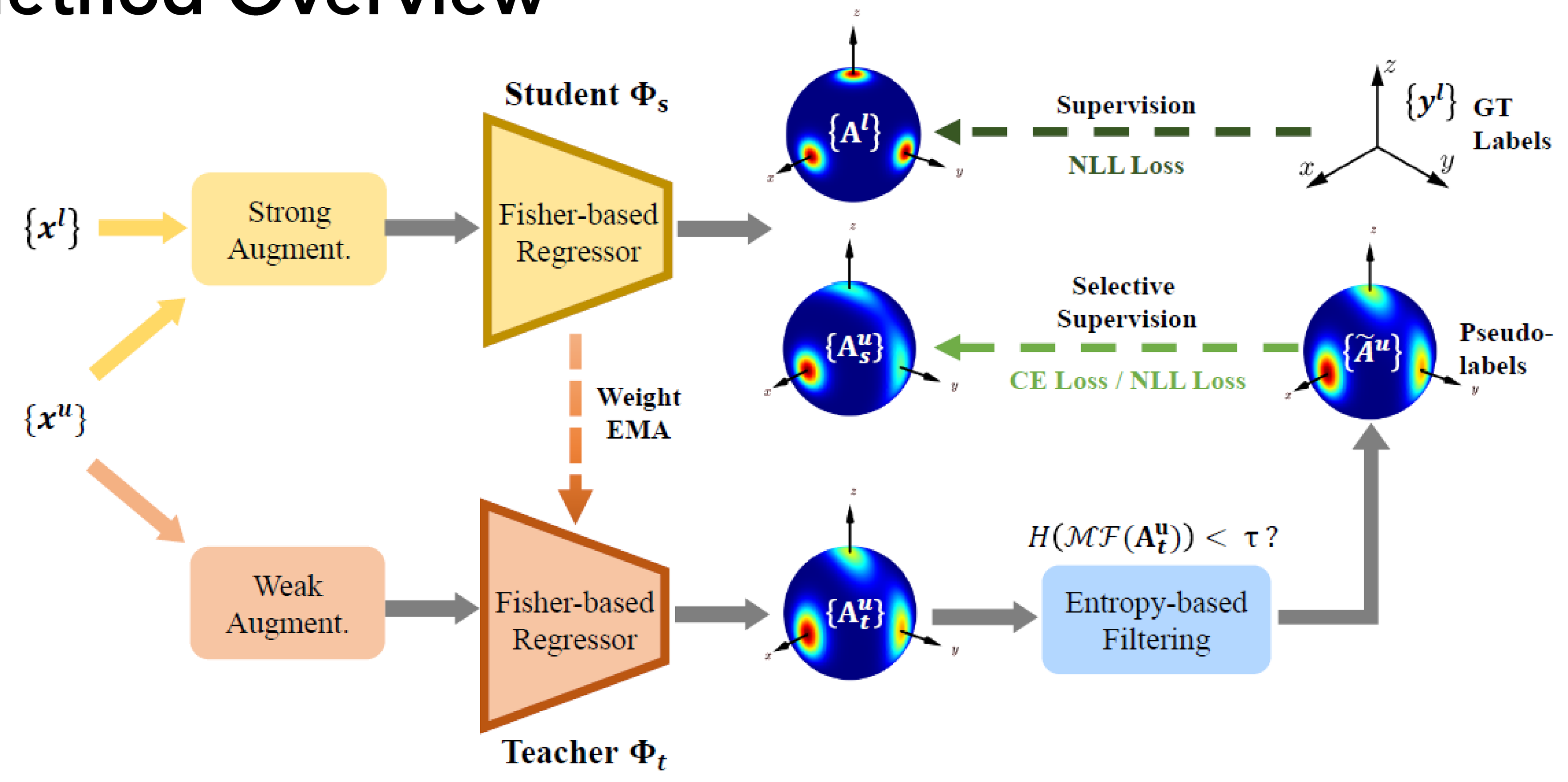




$$\mathcal{B}(\mathbf{q};\Lambda,\mathbf{V})=\frac{1}{F(\Lambda)}\exp\left(\mathbf{q}^T\mathbf{V}\Lambda\mathbf{V}^T\mathbf{q}\right) \qquad \mathcal{MF}(\mathbf{R};\mathbf{A})=\frac{1}{F(\mathbf{A})}\exp\left(\operatorname{tr}\left(\mathbf{A}^T\mathbf{R}\right)\right)$$

- Probabilistic modeling of rotation is the correct way to model the *uncertainty* of rotation regression.
- For one matrix Fisher distribution, an equivalent Bingham distribution exists and satisfy $p_F = 2\pi^2 p_B$.
- We choose matrix Fisher due to the continuity of the rotation matrices.

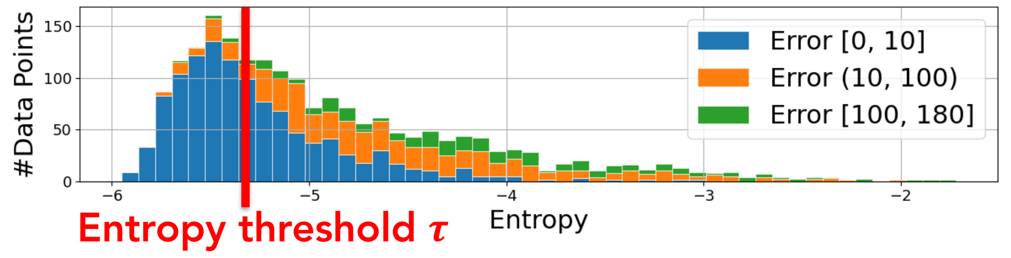
Method Overview



- We adopt teacher-student framework with pseudo-label filtering, where **pseudo label is a matrix Fisher distribution on SO(3)**.
- Key differences: 1) we use analytical cross entropy for supervising the errors between two matrix Fisher distributions.
  2) we utilize the **distribution entropy as rotation uncertainty** to perform pseudo-label filtering.

- **Entropy-based pseudo-label filtering**

Entropy is highly correlated with errors!



- **Loss**

Overall loss $\quad L = L_l\left(\boldsymbol{x}^l,\boldsymbol{y}^l\right)+\lambda_u L_u\left(\boldsymbol{x}^u\right)$

Supervised loss: Negative log likelihood

$$L_l\left(\boldsymbol{x}^l,\boldsymbol{y}^l\right)=-\log\left(\mathcal{MF}\left(\boldsymbol{y}^l;\mathbf{A}^l\right)\right)$$

Unsupervised loss with entropy-based filtering

$$L_u\left(\boldsymbol{x}^u\right)=\mathbb{1}\left(H(p_t)\le\tau\right)L\left(p_t,p_s\right)$$

Unsupervised CE loss $\quad L^{\mathrm{CE}}\left(\delta(\boldsymbol{y}_t^u),p_s\right)=H\left(\delta(\boldsymbol{y}_t^u),p_s\right)$

Unsupervised NLL loss $\quad L^{\mathrm{NLL}}\left(p_t,p_s\right)=-\log p_s(\boldsymbol{y}_t^u)$

Experiment

- Superior performance over supervised and other semi-supervised baselines
- Few annotations required

Results on ModelNet10-SO(3) under different ratios of labeled data

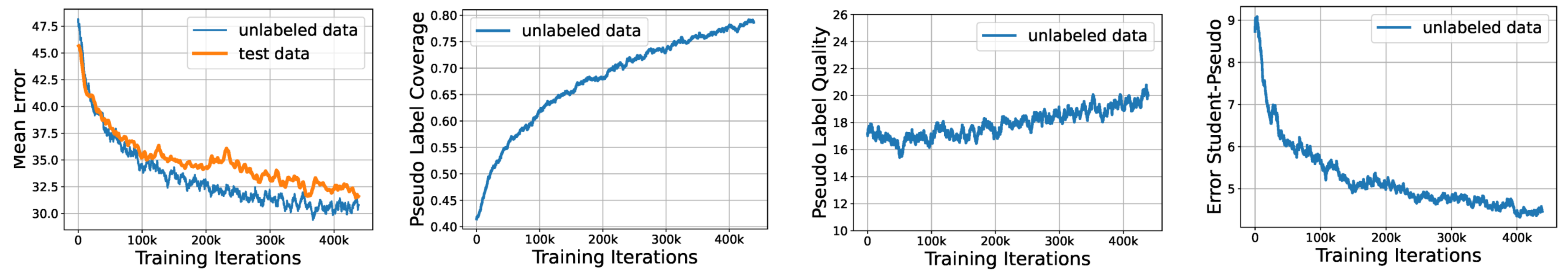
Category	Method	5%		10%	
		Mean↓	Med.↓	Mean↓	Med.↓
Sofa	Sup.-L1 [29]	44.64	11.42	32.65	9.03
	Sup.-Fisher [35]	45.19	13.16	32.92	8.83
	SSL-L1-Consist.	36.86	8.65	25.94	6.81
	SSL-FisherMatch	**32.02**	**7.78**	**21.29**	**5.25**
	Full Sup.	18.62	5.77	18.62	5.77
Chair	Sup.-L1 [29]	40.41	16.09	29.02	10.64
	Sup.-Fisher [35]	39.34	16.79	28.58	10.84
	SSL-L1-Consist.	31.20	11.29	23.59	8.10
	SSL-FisherMatch	**26.69**	**9.42**	**20.06**	**7.44**
	Full Sup.	17.38	6.78	17.38	6.78

Results on Pascal3D+ dataset with few labeled images

Method	7		20		50	
	Med.↓	$Acc_{30°}$↑	Med.↓	$Acc_{30°}$↑	Med.↓	$Acc_{30°}$↑
Res50-Gene	39.1	36.1	26.3	45.2	20.2	54.6
Res50-Spec	46.5	29.6	29.4	42.8	23.0	50.4
StarMap [59]	49.6	30.7	46.4	35.6	27.9	53.8
NeMo [45]	60.0	38.4	33.3	51.7	22.1	69.3
NVSM [46]	37.5	53.8	28.7	61.7	24.2	65.6
FisherMatch	**28.3**	**56.8**	**23.8**	**63.6**	**16.1**	**75.7**
Full Sup.	8.1	89.6	8.1	89.6	8.1	89.6

Analysis

Training process



- CE loss performs the better with broader compatibility, and the NLL loss encourages a higher confidence of the network.
- NLL loss is a sharpened version of CE loss: $L^{\mathrm{CE}}\left(\mathrm{Dirac}(p_t),p_s\right)=L^{\mathrm{NLL}}\left(p_t,p_s\right)$