Meta Soft Label Generation for Noisy Labels

Görkem Algan
Ilkay Ulusoy

Outline

- Motivation and Problem Definition
- Preliminary Knowledge on Meta-Learning
- Proposed Algorithm
- Experimental Results

Outline

- **Motivation and Problem Definition**
- Preliminary Knowledge on Meta-Learning
- Proposed Algorithm
- Experimental Results

Motivation and Problem Definition

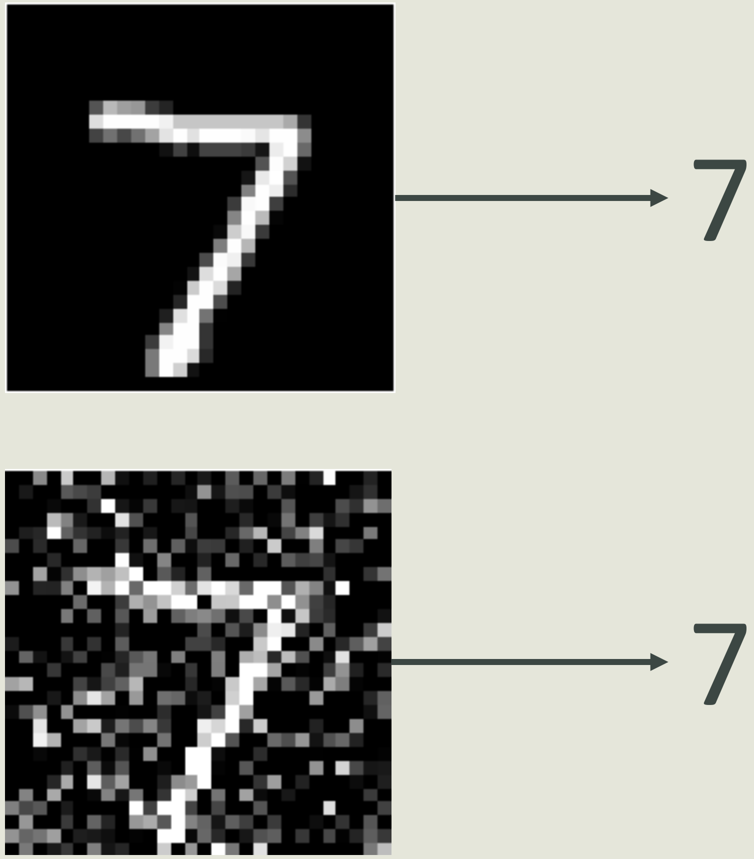
- Reasons behind recent success of deep learning
  - Computational power
  - Large datasets

- Problems with large datasets
  - Hard to obtain clean data
  - Hard to label whole dataset
  - Hard to be sure about labeler accuracy

- As a result “**noisy datasets**”

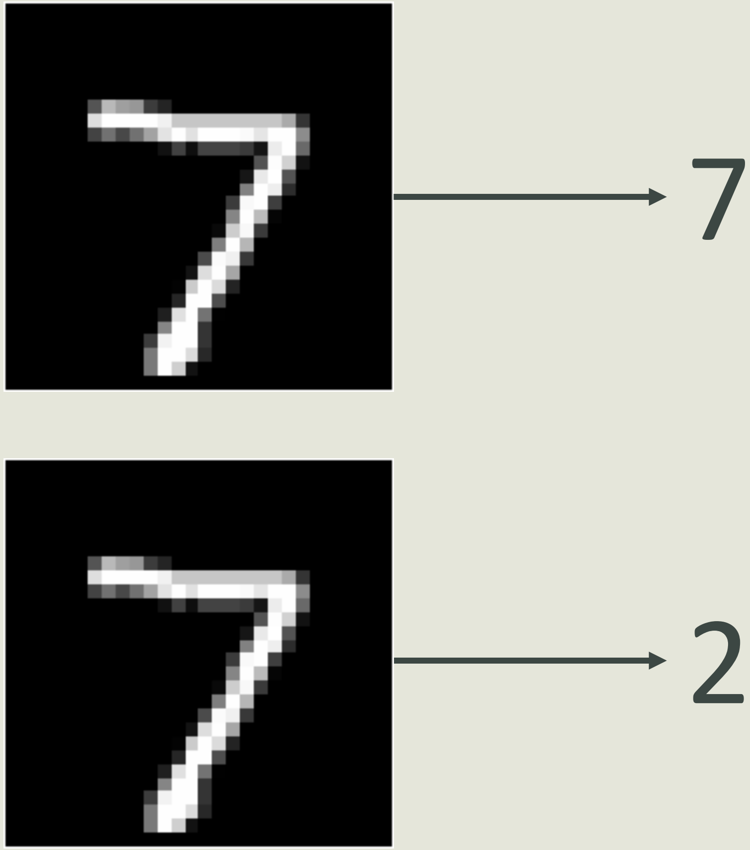
Motivation and Problem Definition

- Two types of noise
  - Feature noise
  - Label noise



Motivation and Problem Definition

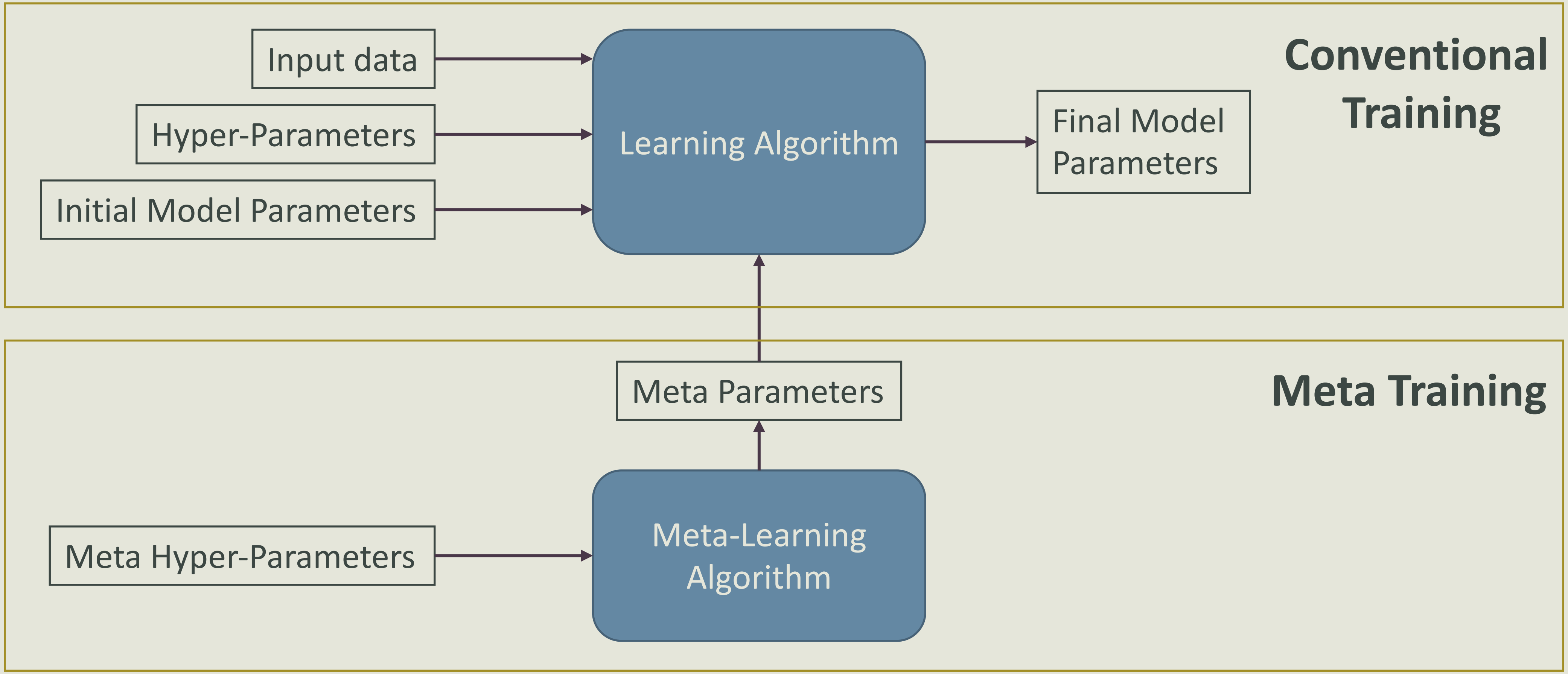
- Two types of noise
  - Feature noise
  - **Label noise**



Outline

- Motivation and Problem Definition
- **Preliminary Knowledge on Meta-Learning**
- Proposed Algorithm
- Experimental Results

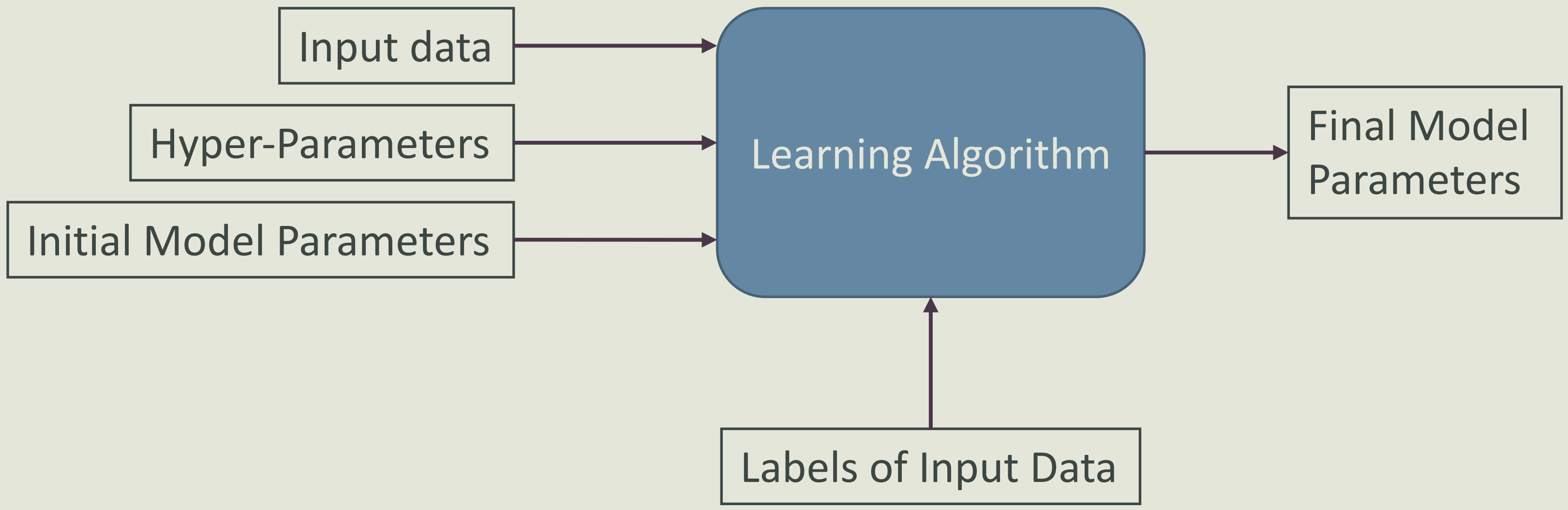
Meta Learning



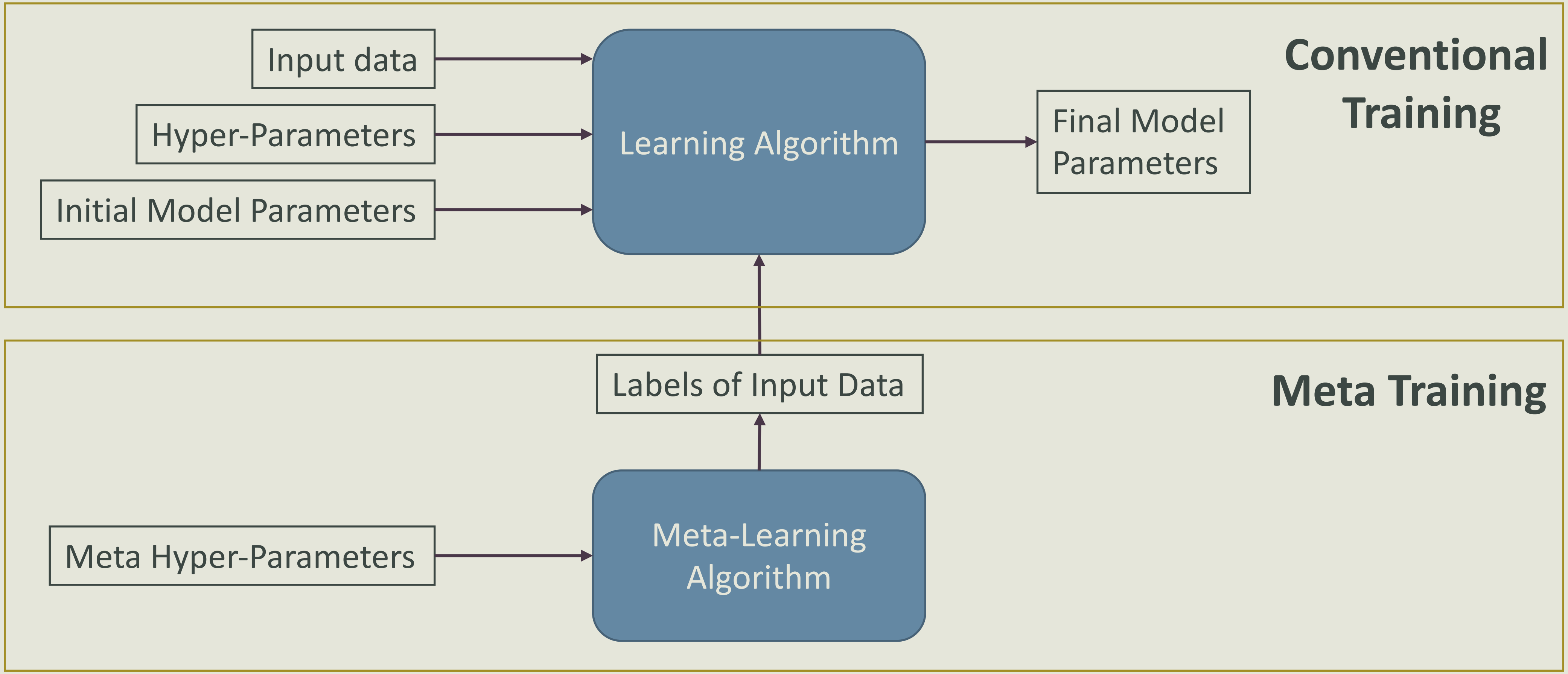
Outline

- Motivation and Problem Definition
- Preliminary Knowledge on Meta-Learning
- **Proposed Algorithm**
- Experimental Results

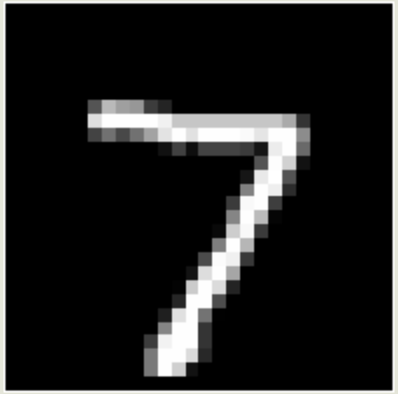
Meta Soft Label Generation for Noisy Labels



Meta Soft Label Generation for Noisy Labels



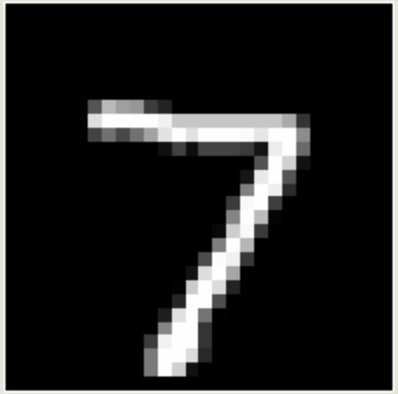
Hard Label vs Soft Label



Hard Label	
0:	0
1:	0
2:	0
3:	0
4:	0
5:	0
6:	0
7:	**1**
8:	0
9:	0

Soft Label	
0:	0
1:	0.3
2:	0.1
3:	0
4:	0
5:	0
6:	0
7:	**0.6**
8:	0
9:	0

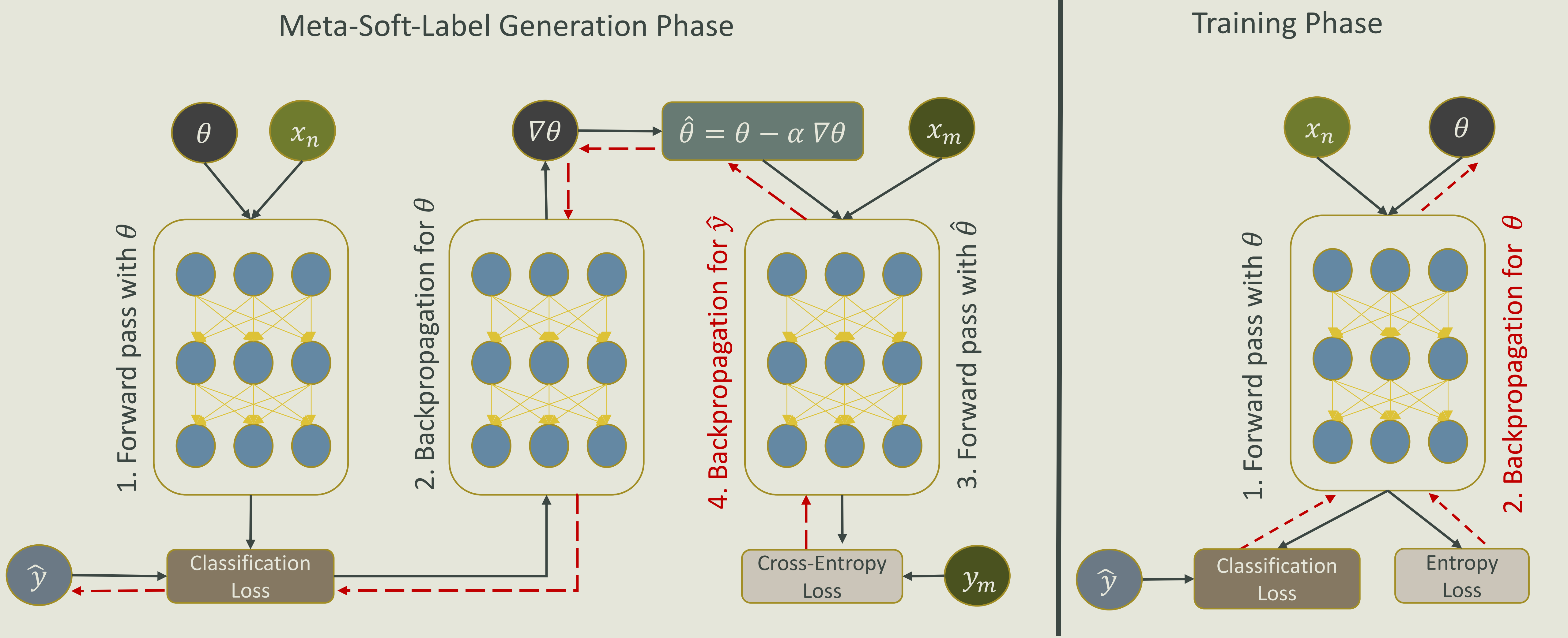
Hard Label vs Soft Label



Hard Label	
0:	0
1:	0
2:	0
3:	0
4:	0
5:	0
6:	0
7:	**1**
8:	0
9:	0

Soft Label	
0:	0
1:	**0.3**
2:	**0.1**
3:	0
4:	0
5:	0
6:	0
7:	**0.6**
8:	0
9:	0

Meta Soft Label Generation for Noisy Labels



Loss functions

- Classification loss: KL($f_\theta(x_i)||y_i$) = $-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} f_\theta^j(x_i)\log(\frac{f_\theta^j(x_i)}{y_i^j})$
- Entropy loss: $-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} f_\theta^j(x_i)\log(f_\theta^j(x_i))$

Meta Soft Label Generation for Noisy Labels

- Two stage learning:
  1. Warm-up training with noisy labels
  2. Learning with proposed algorithm

- Warm-up training for
  - Deep networks are highly noise robust in initial epochs
  - Random network predictions are bad for meta-objective
    - Unstabilised meta-gradients

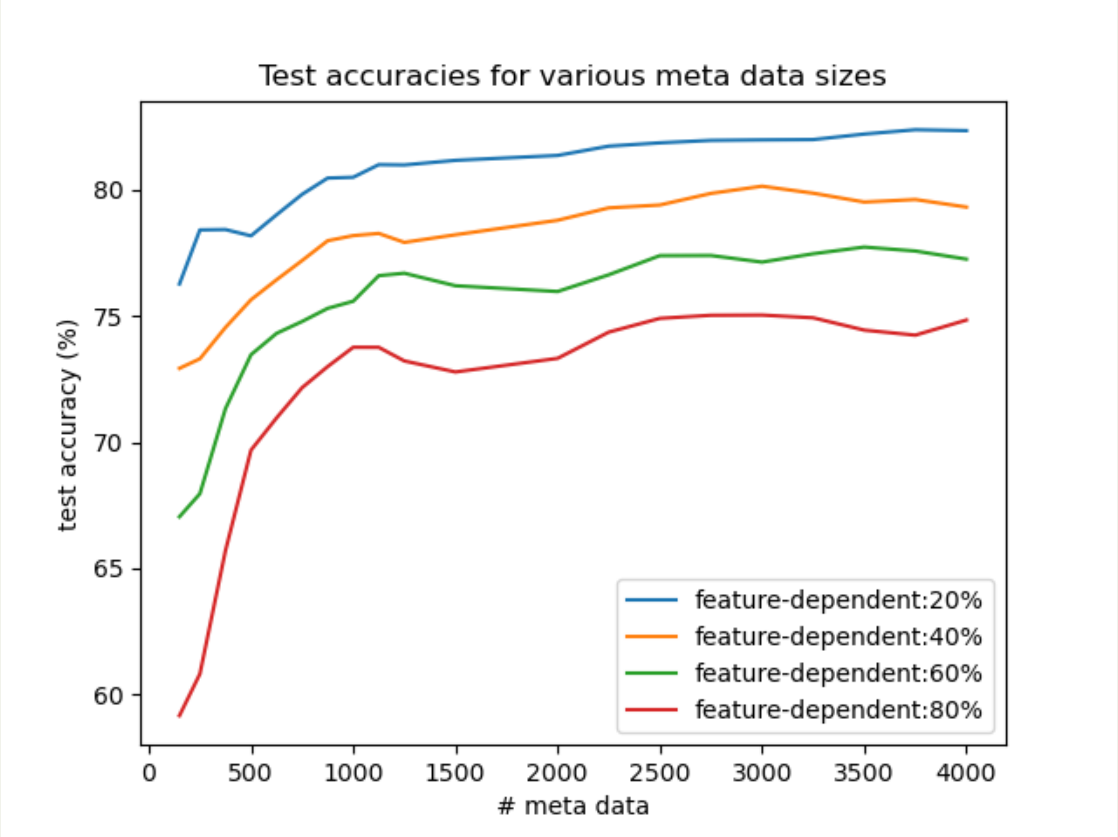
Outline

- Motivation and Problem Definition
- Preliminary Knowledge on Meta-Learning
- Proposed Algorithm
- **Experimental Results**

CIFAR10 Results

Noise type	Uniform				Feature-dependent			
Noise ratio (%)	**20**	**40**	**60**	**80**	**20**	**40**	**60**	**80**
Cross Entropy	82.69	76.84	66.46	38.04	81.21	71.46	69.19	23.89
Symmetric CE	82.72	79.79	74.09	54.56	76.21	67.76	fail	fail
Generalized CE	84.62	**81.98**	**74.48**	44.36	81.21	71.80	66.56	fail
Bootstrap	82.51	76.97	66.13	38.41	81.24	71.63	69.74	23.25
Co-teaching	85.96	80.24	70.38	41.22	81.19	72.47	67.67	18.66
Forward Loss	83.31	80.25	71.34	28.77	77.60	69.21	39.23	fail
Joint Opt.	83.74	78.75	68.17	39.22	81.61	74.03	72,15	44.15
PENCIL	**83.34**	79.27	71.41	46.57	81.62	75.08	69.24	fail
MLNT	83.20	78.14	66.34	40.80	82.46	72.52	70.12	Faul
Meta-Weight	84.12	80.68	71.78	46.71	81.06	71.50	67.50	22.28
MSLG	83.48	78.82	72.92	**56.26**	**82.62**	**79.30**	**77.33**	**74.87**

Test Accuracies for Different Numbers of Meta-Data



Clothing1M Results

Method	Accuracy	Method	Accuracy
Generalized CE	67.85	Joint Optimization	72.16
Bootsrap	69.35	MLNT	73.47
Co-Teaching	69.63	PENCIL	73.49
Forward Loss	70.94	Meta-Weight Net	73.72
Symmetric CE	71.02	**MSLG**	**76.02**

Food101N Results

Method	Accuracy	Method	Accuracy
Generalized CE	71.60	Bootsrap	78.03
Meta-Weight Net	76.12	PENCIL	78.26
Joint Optimization	76.18	Co-Teaching	78.95
	MSLG: 79.06		

Thank You for Listening!