

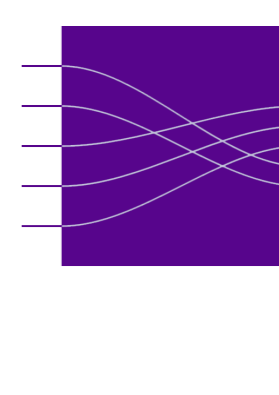
Few-Shot Continual Learning for Audio Classification

Yu Wang¹, Nicholas J. Bryan², Mark Cartwright¹, Juan Pablo Bello¹, Justin Salamon²

¹Music and Audio Research Laboratory, New York University ²Adobe Research

y-wang.weebly.com

wangyu@nyu.edu



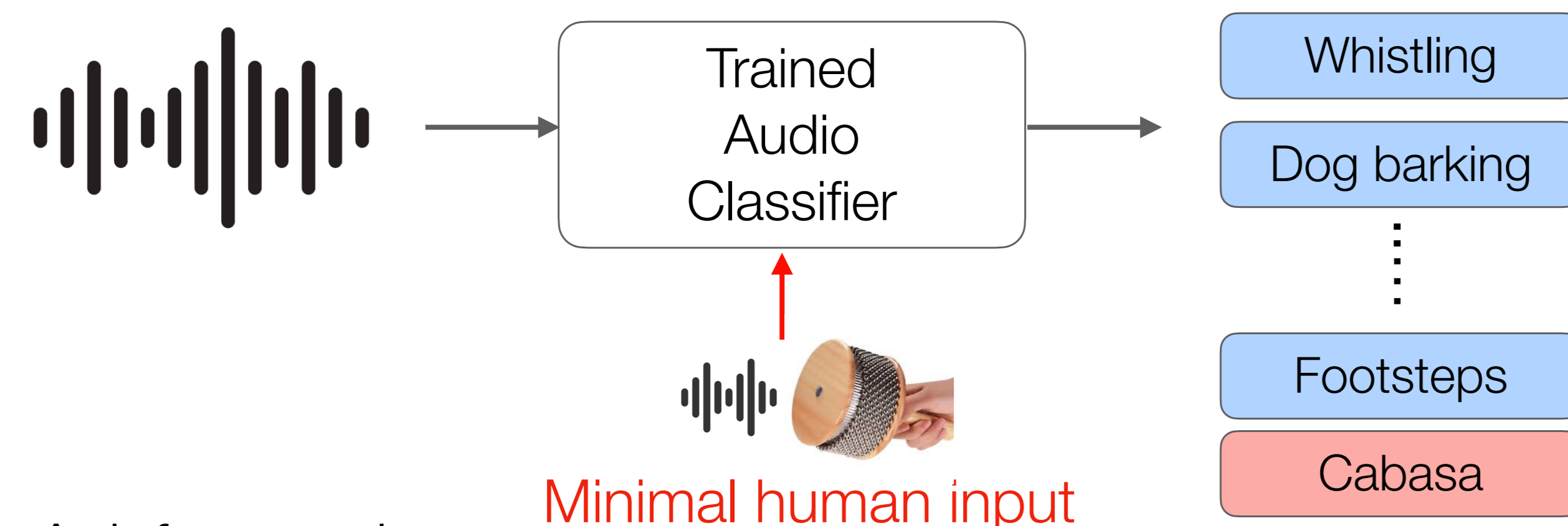
MARL



Adobe



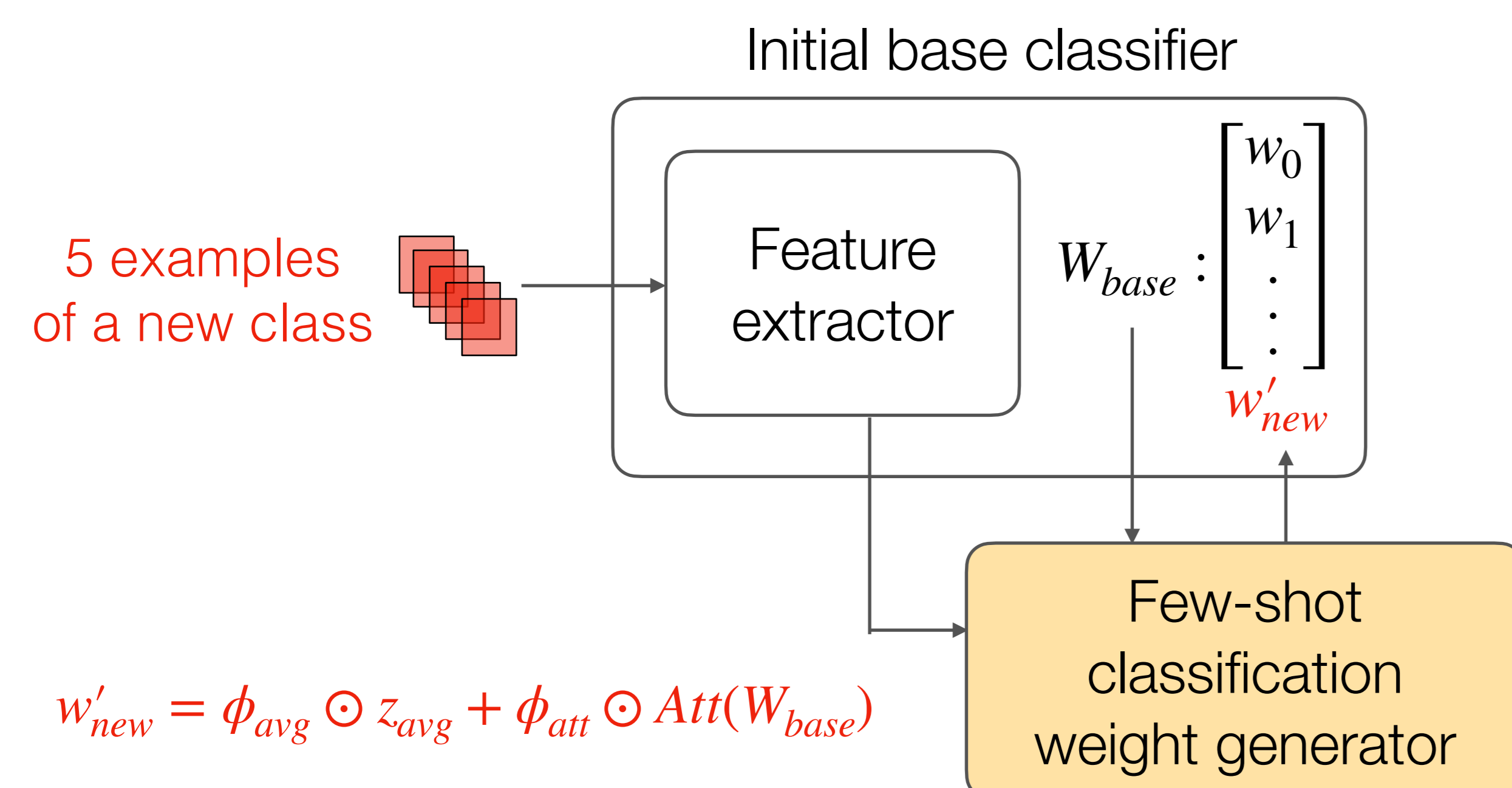
1. Proposed Framework



- At inference time:

- Add a new class on top of existing sound vocabulary
- Based on few labeled data (in this work: 5)

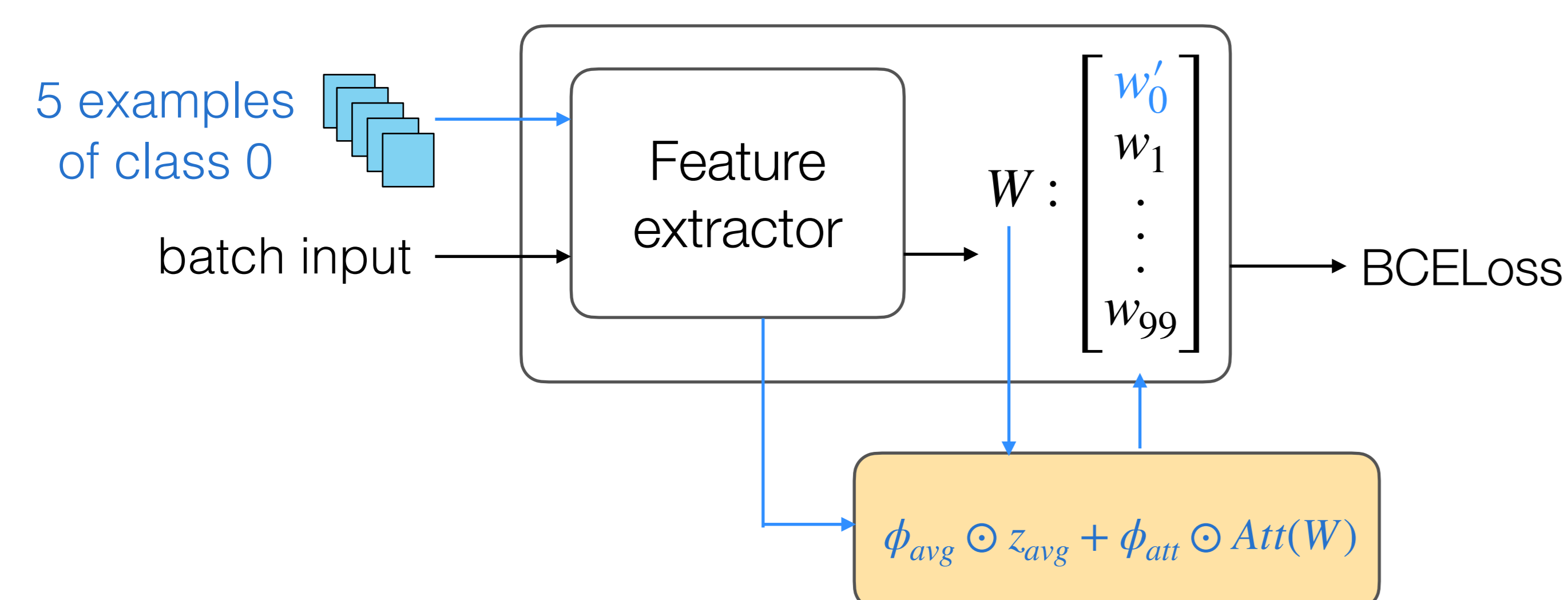
2. Dynamic Few-Shot Learning (DFSL)



3. Training the Classification Weight Generator

- In a training iteration:

- Randomly choose 5 base classes as “pseudo” novel classes
- Generate 5 new classification weight vectors to replace the original ones
- Compute loss based on the updated classification weight matrix



4. Experiment: Datasets

Dataset	AudioSet subset	ESC-50
# Examples	~335k	2000
# Base classes	100	30
# Novel Classes	20	10
Polyphony	Polyphonic	Monophonic
Label Quality	Potentially incomplete	Perfect

- AudioSet subset: leaf classes, annotation quality > 80%

5. Experiment: Training & Evaluation Setup

- PANN model as feature extractor

Log-mel spectrogram (64, 1000)

(64 (3x3) Conv, BN, ReLU) x2

Mean Pooling (2x2)

(128 (3x3) Conv, BN, ReLU) x2

Mean Pooling (2x2)

(256 (3x3) Conv, BN, ReLU) x2

Mean Pooling (2x2)

(512 (3x3) Conv, BN, ReLU) x2

Mean Pooling (2x2)

(1024 (3x3) Conv, BN, ReLU) x2

Mean Pooling (2x2)

(2048 (3x3) Conv, BN, ReLU) x2

Global Pooling

2048 FC, ReLU

- Training:

- Data: base classes
- Loss: BCELoss
- 2-stage training
 - Freeze the feature extractor after first training stage
- Use pretrained feature extractor for ESC-50

- Evaluation:

- 5 labeled examples available for each novel class
- Test data: base & novel classes
- Joint label space

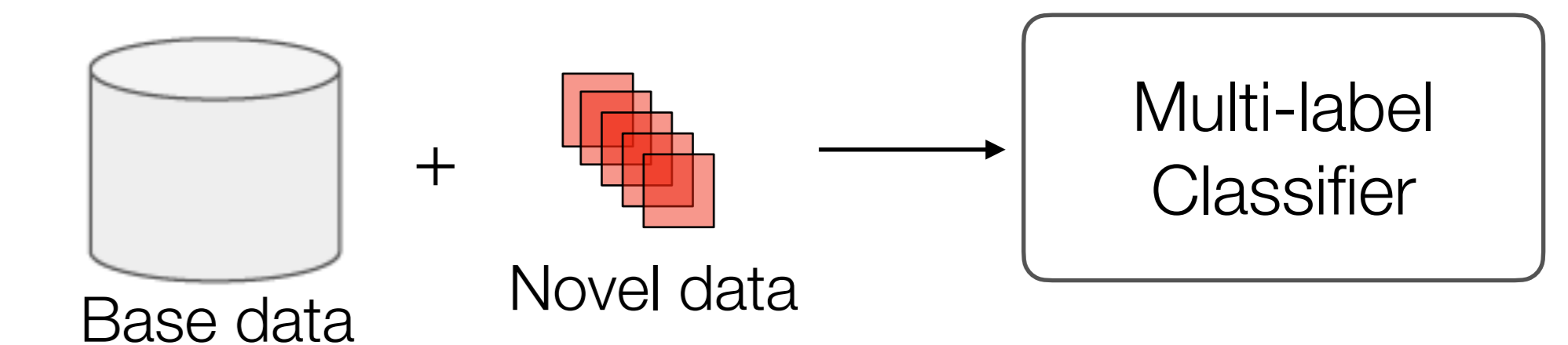
7. Experiment: Evaluation Results

Methods	F-measure	
	Base classes (100)	Novel Classes (20)
Base Clf.	0.50	-
DFSL (Ours)	0.48	0.21
Retrain	0.44	0.06
Prototypical Network	(Requires threshold tuning)	
Base Clf.+Prototype	0.50	0.08

6. Experiment: Baselines

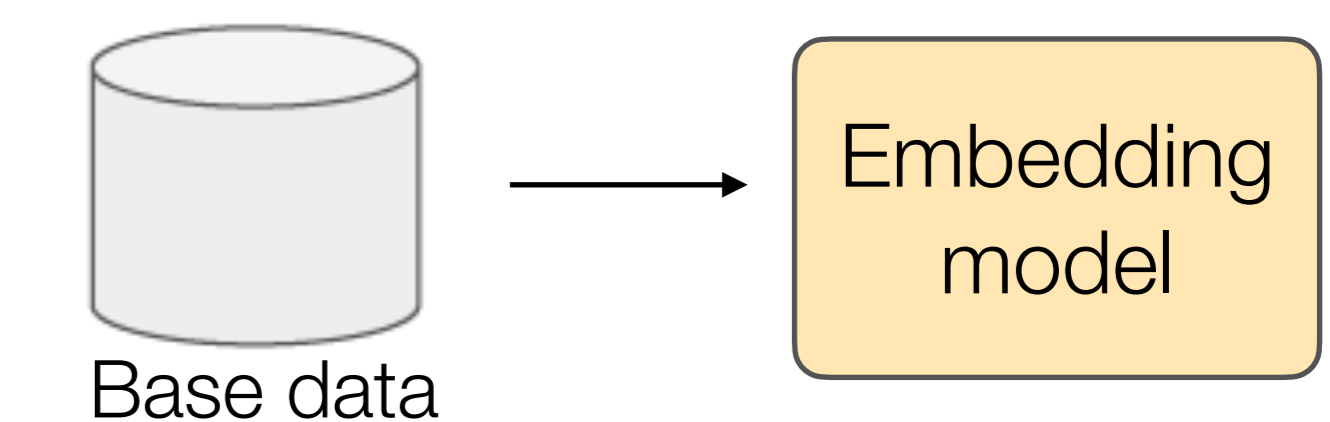
- Base classes: many labeled data
- Novel classes: 5 labeled data per class

1. Retrain (standard supervised learning)

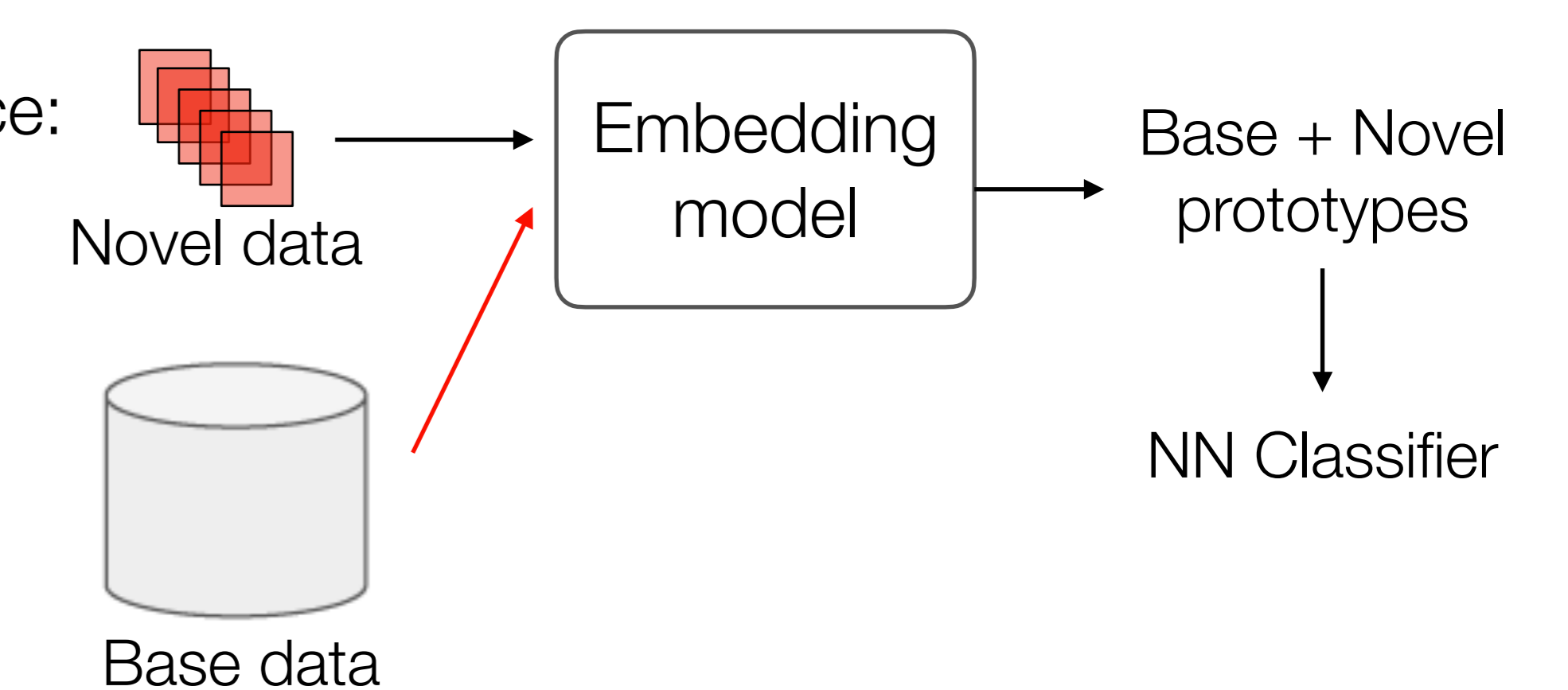


2. Prototypical Network (few-shot learning)

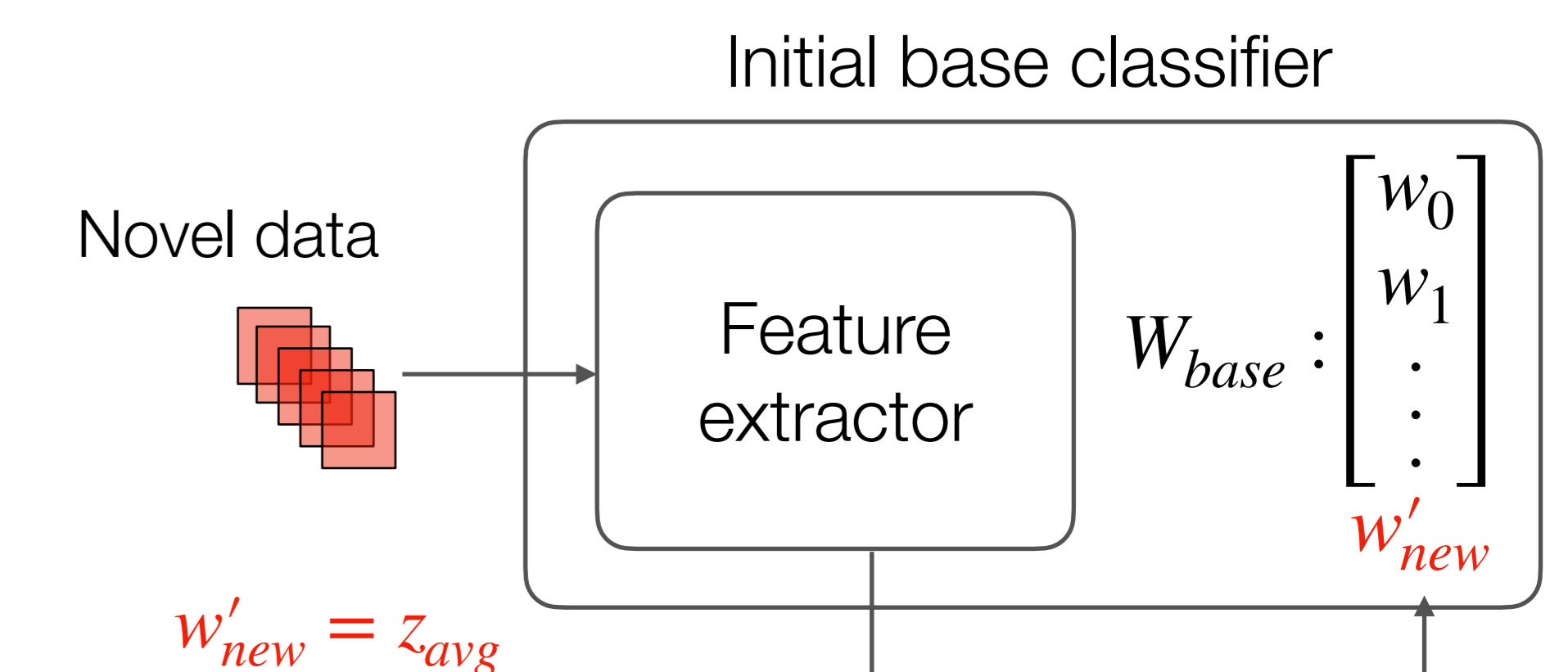
- Training:



- Inference:



3. Initial base classifier + prototype



Methods	F-measure	
	Base classes (30)	Novel Classes (10)
Base Clf.	0.62	-
DFSL (Ours)	0.59	0.53
Retrain	0.59	0.52
Prototypical Network	(Requires threshold tuning)	
Base Clf.+Prototype	0.62	0.26