



E²CM: Early Exit via Class Means for Efficient Supervised and Unsupervised Learning



Alperen Gormez[†], Venkat R. Dasari[§], Erdem Koyuncu[†]

[†]University of Illinois Chicago

[§]CCDC US Army Research Laboratory

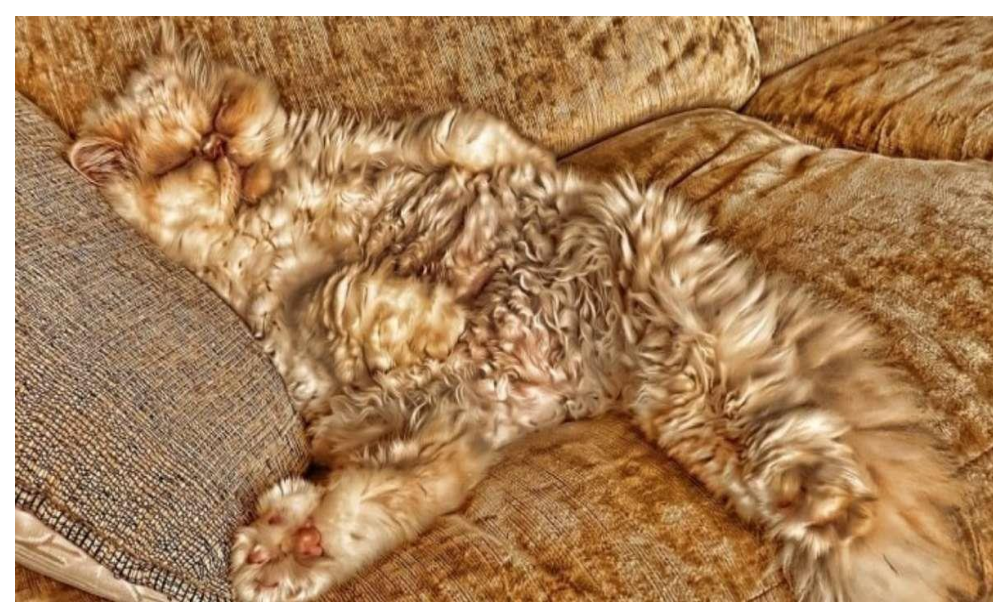
agorme2@uic.edu, venkateswara.r.dasari.civ@army.mil, ekoyuncu@uic.edu

Motivation

- Idea of early exit: Not all examples need to traverse the entire neural network, simple examples should exit early from the network.

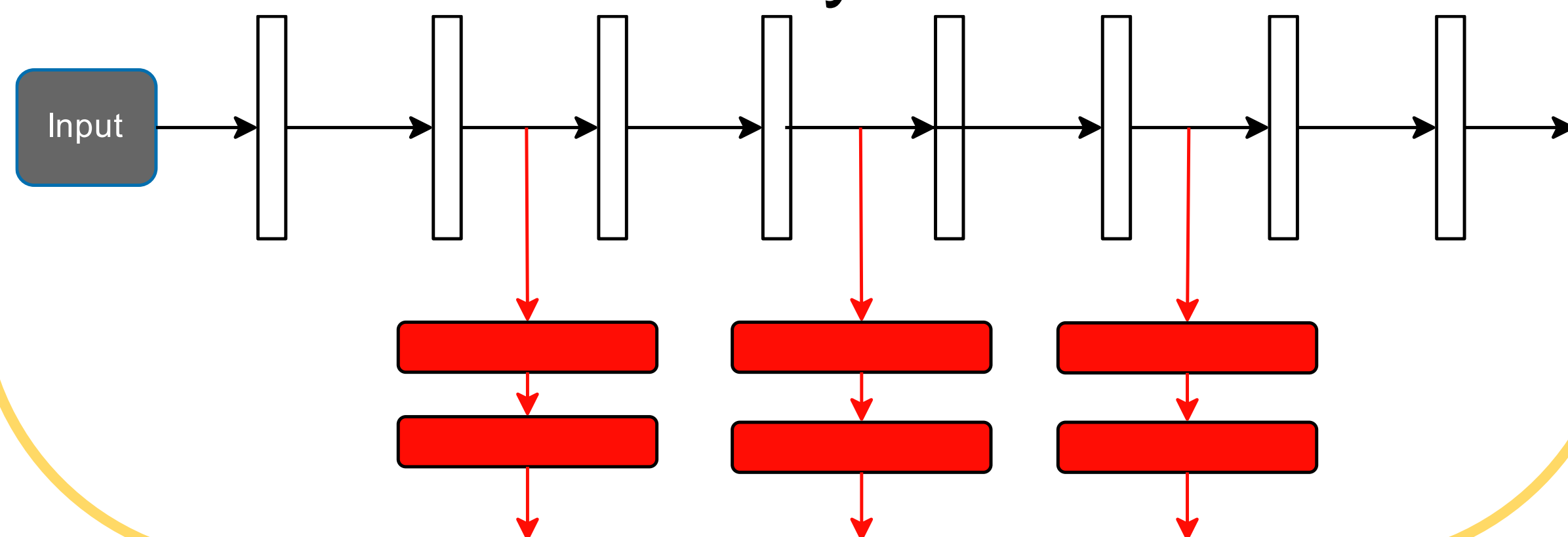


Classification Difficulty: **Easy**
Early Exit: **Yes**



Classification Difficulty: **Hard**
Early Exit: **No**

- Existing early exit methods are not suitable for the **edge**, because they:
 - Add** exit layers that require gradient based training to the base network.
 - Train** exit layers jointly with the base model.
 - Need** additional hyperparameters such as the location, number and size of the exit layers.

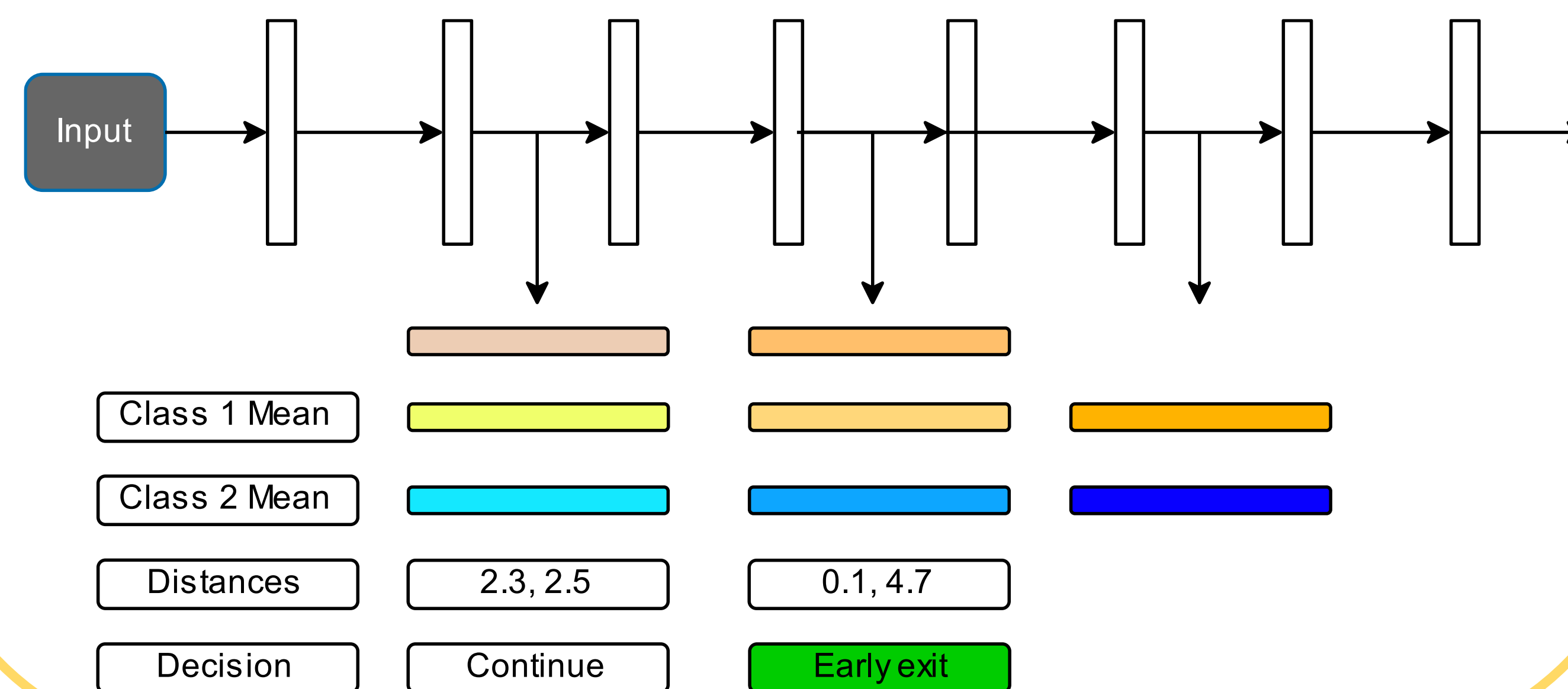


Our Contribution

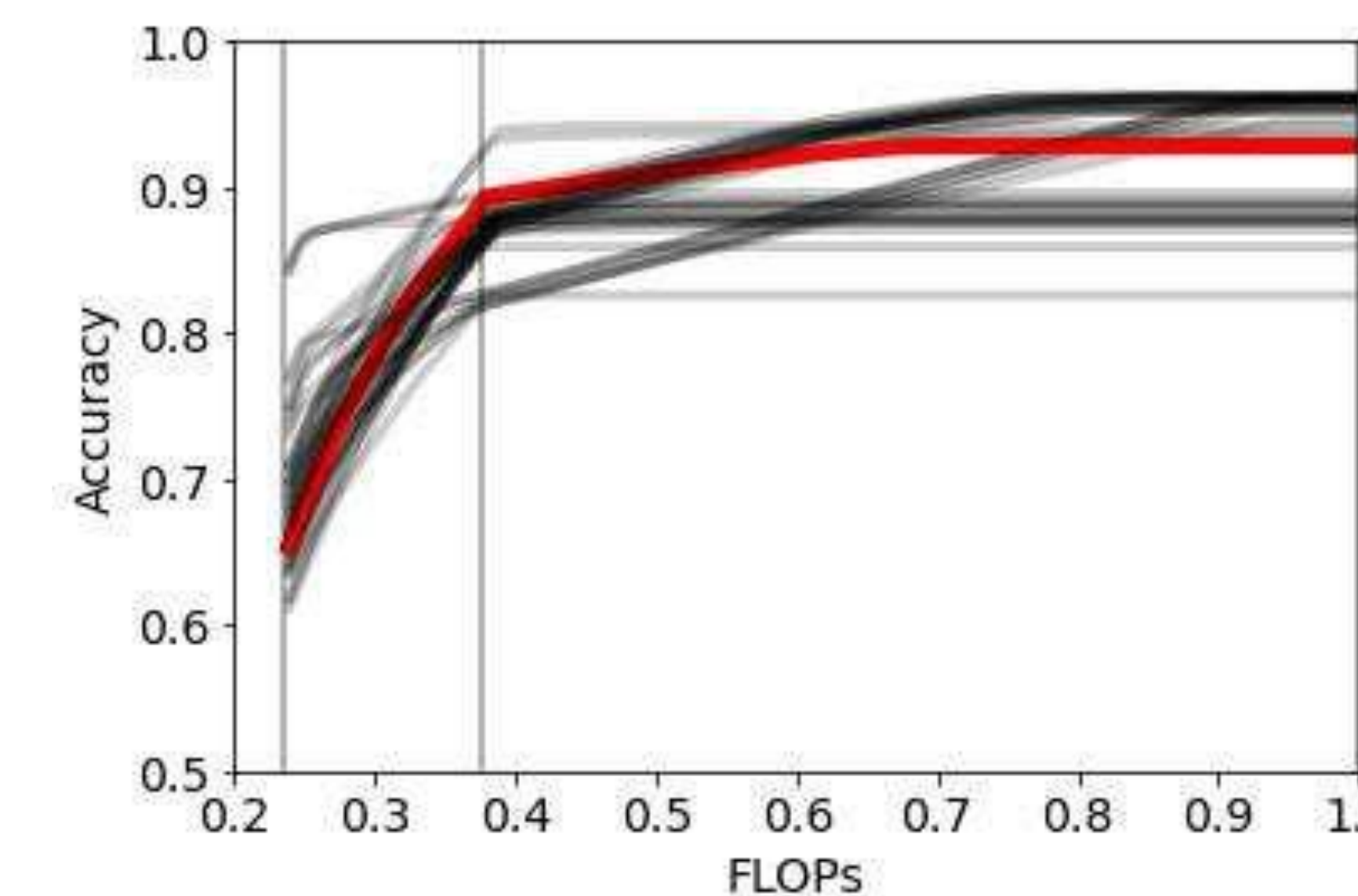
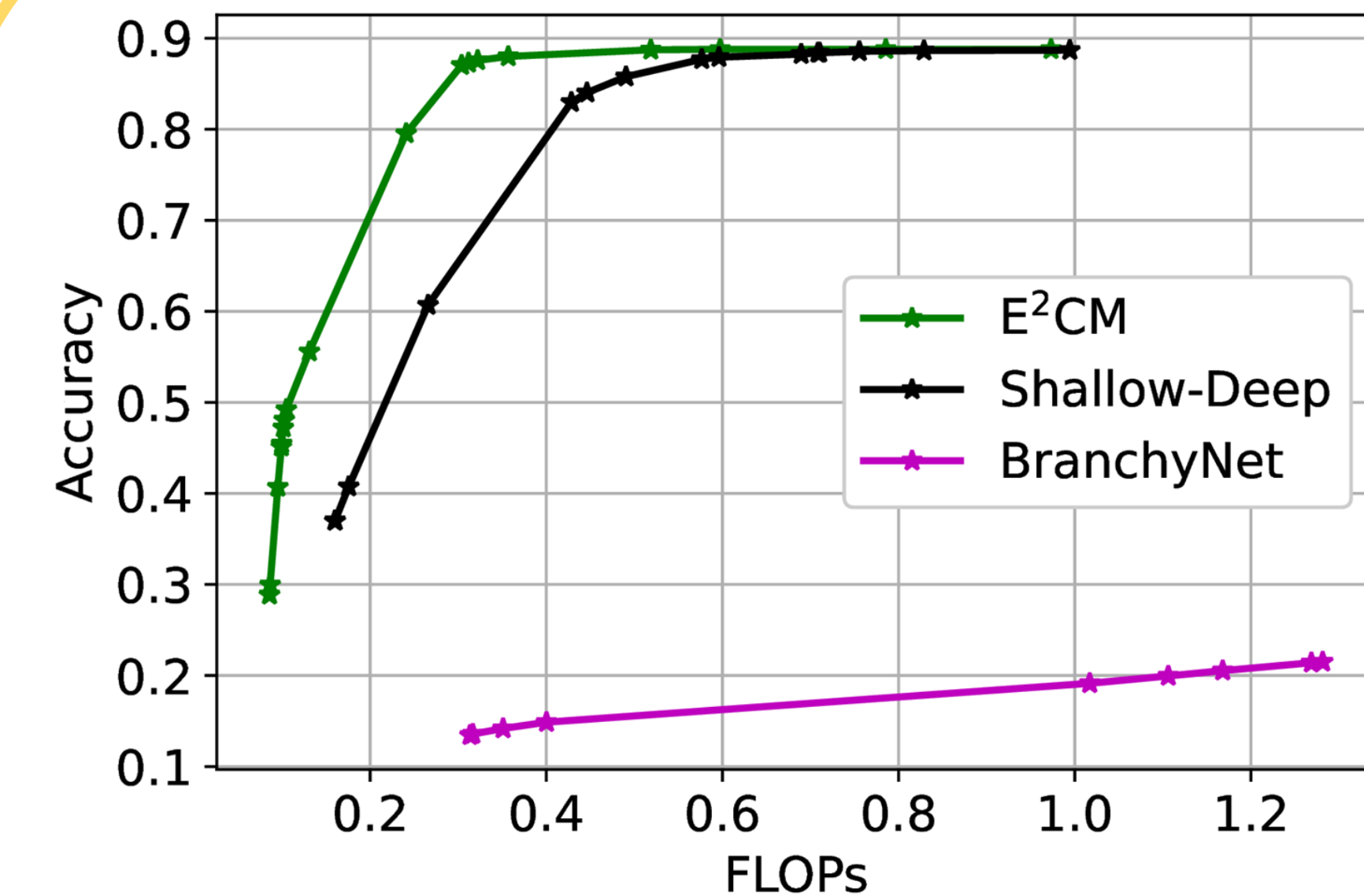
- E²CM**: A simple, lightweight early exit method.
- E²CM does not add trainable layers, therefore it **does not modify** the base network.
- E²CM **does not need** gradient based training.
- E²CM **does not need** additional hyperparameters.

E²CM Algorithm

- E²CM calculates class means by **averaging** feature vectors at each layer for each class using training set examples.
- During prediction, E²CM calculates the **distances** between the feature vector and the class means. If the feature vector is close enough to a class mean, the sample exits early and that class is predicted. Otherwise, the sample moves forward to the next exit point.



Results



Model, Dataset	Method	Accuracy			
		$\phi=0.15$	$\phi=0.20$	$\phi=0.25$	$\phi=0.30$
ResNet-152, CIFAR-10	E ² CM+SDN	82.4%	86.4%	88.5%	88.8%
	SDN	80%	86%	88.4%	88.8%
	BWDS	82%	84.1%	86%	87.7%
	BranchyNet	80%	80%	80%	80%

Conclusion

E²CM is a simple and lightweight early exit method that **does not modify** the base network and **does not need** gradient based training. It can be applied to **both supervised and unsupervised** learning tasks.