Scaling Video Analytics on Constrained Edge Nodes

Christopher Canel, Thomas Kim, Giulio Zhou, Conglong Li, Hyeontaek Lim,
David G. Andersen, Michael Kaminsky[†], Subramanya R. Dulloor[‡]

Carnegie Mellon University [†] Intel Labs [‡] ThoughtSpot

April 1, 2019

In smart cities, cameras are ubiquitous

More ***deployments***, and importantly, more ***applications***

Cameras in Hangzhou, Zhejiang Province, China

Traffic monitoring, pedestrian tracking, event detection, etc.

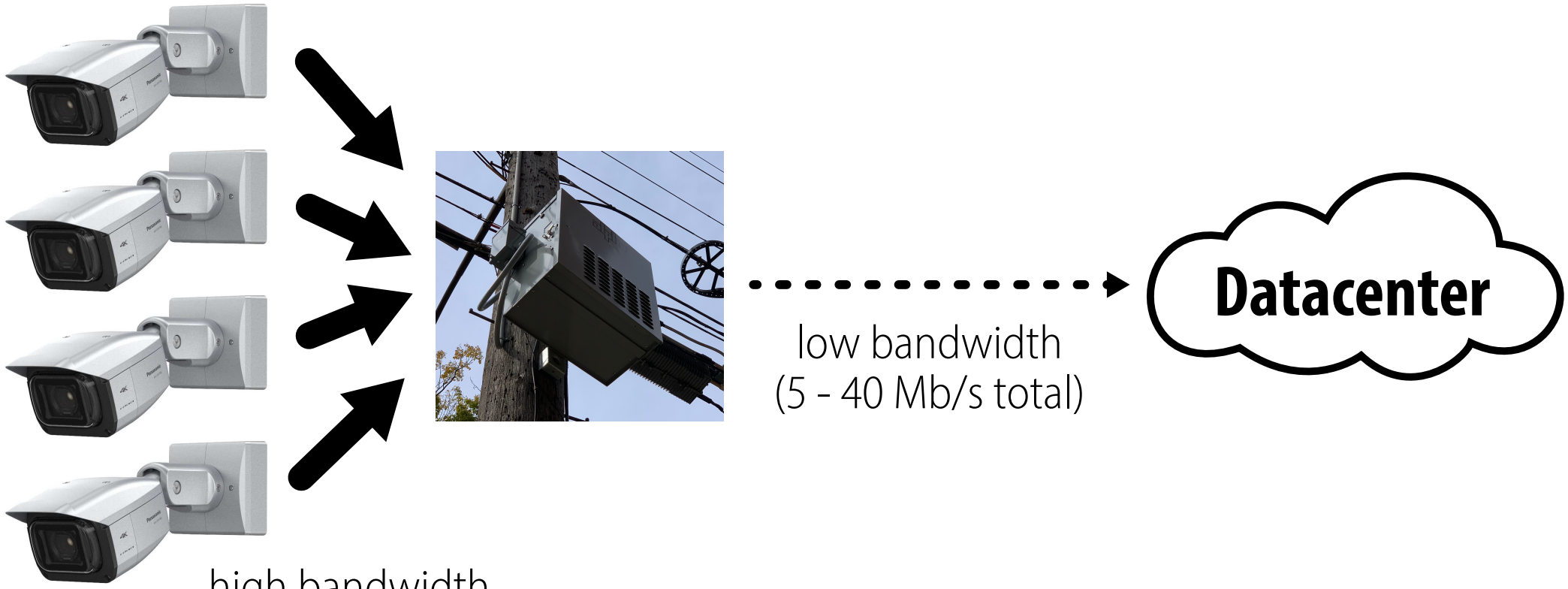
This talk

Bandwidth ***challenges*** in edge-to-cloud camera deployments

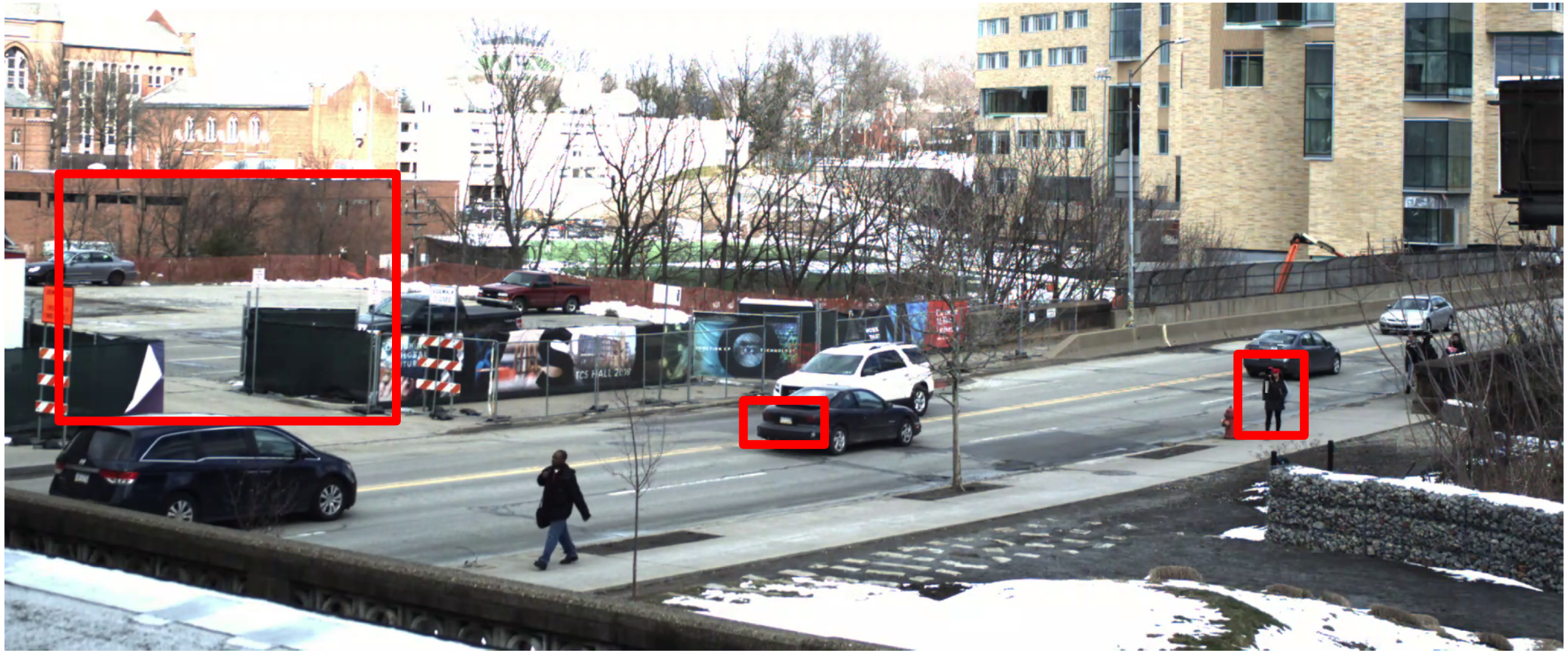
Introduce ***FilterForward***, a new system for efficient video filtering on the edge

Discuss ***microclassifiers***, cheap but accurate application-specific NN filters

Bandwidth is limited



Many apps interested in different content ⇨ Cannot crop to a small ROI



Targeting apps that need subtle details ⇨ Cannot heavily compress the entire stream

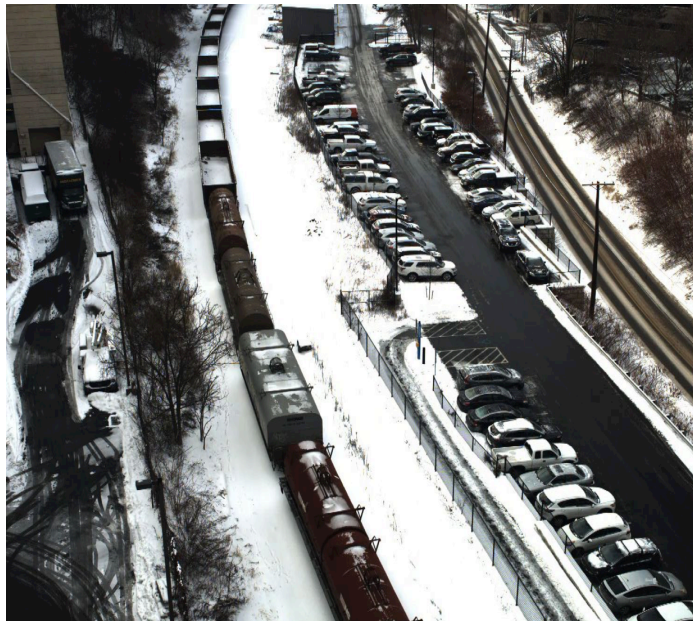
Compression and down-sampling destroy subtle details that applications need

Full frame

7000 Kb/s

100 Kb/s

Target environment: Apps interested in rare events



E.g., specific makes/models of vehicles, particular social interactions, etc.

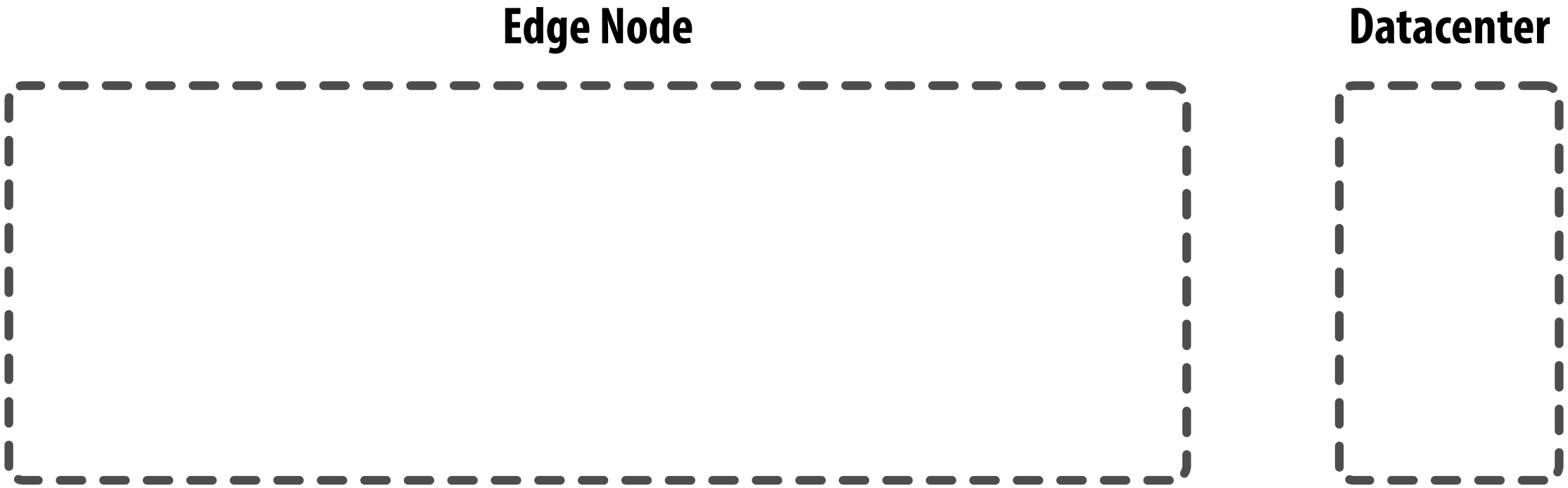
Opportunity to save bandwidth by dropping uninteresting frames

Caveat: Filtering, especially for many apps, requires more compute resources

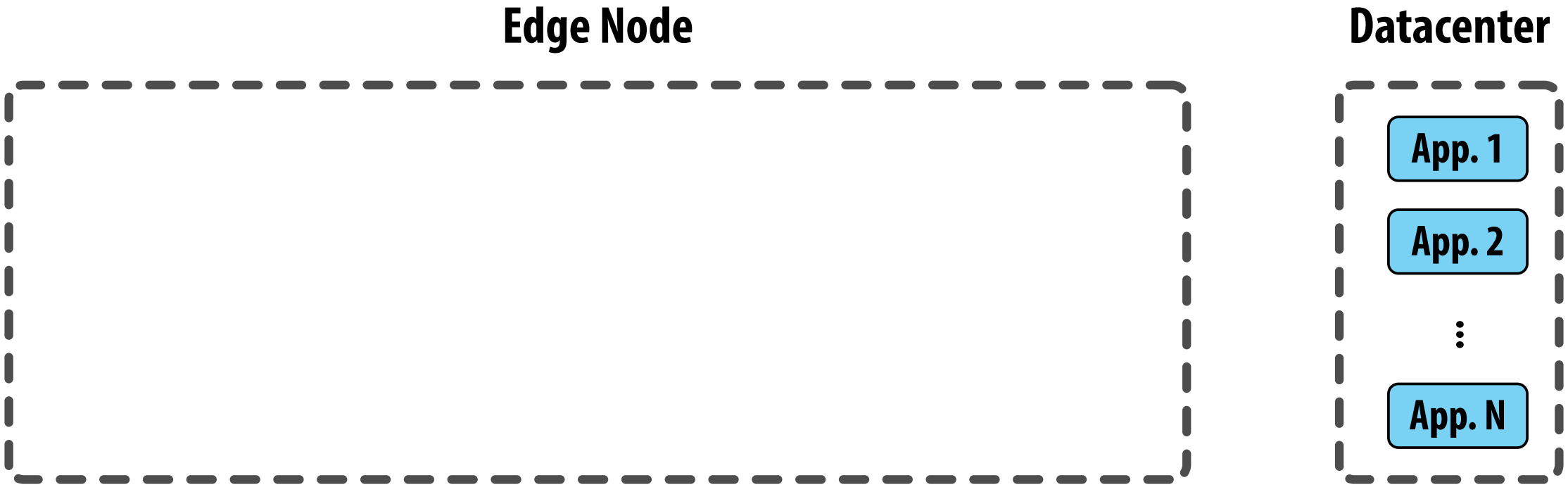
Our work

FilterForward saves bandwidth by filtering relevant data at the edge, while supporting multi-tenancy

FilterForward at a glance

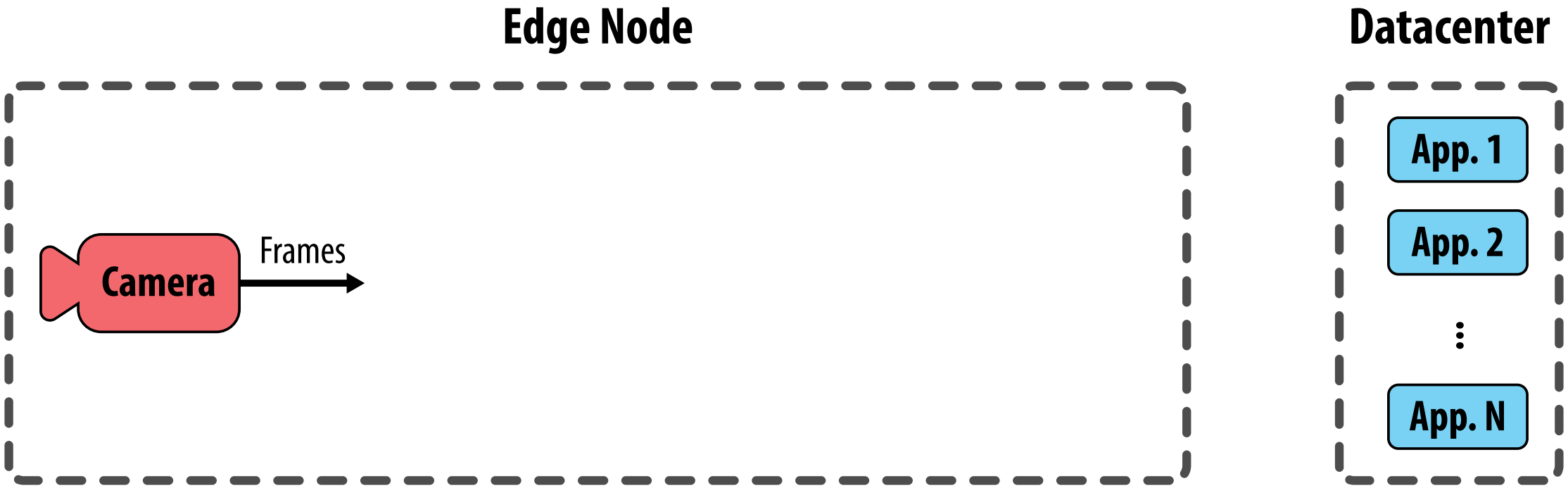


FilterForward at a glance



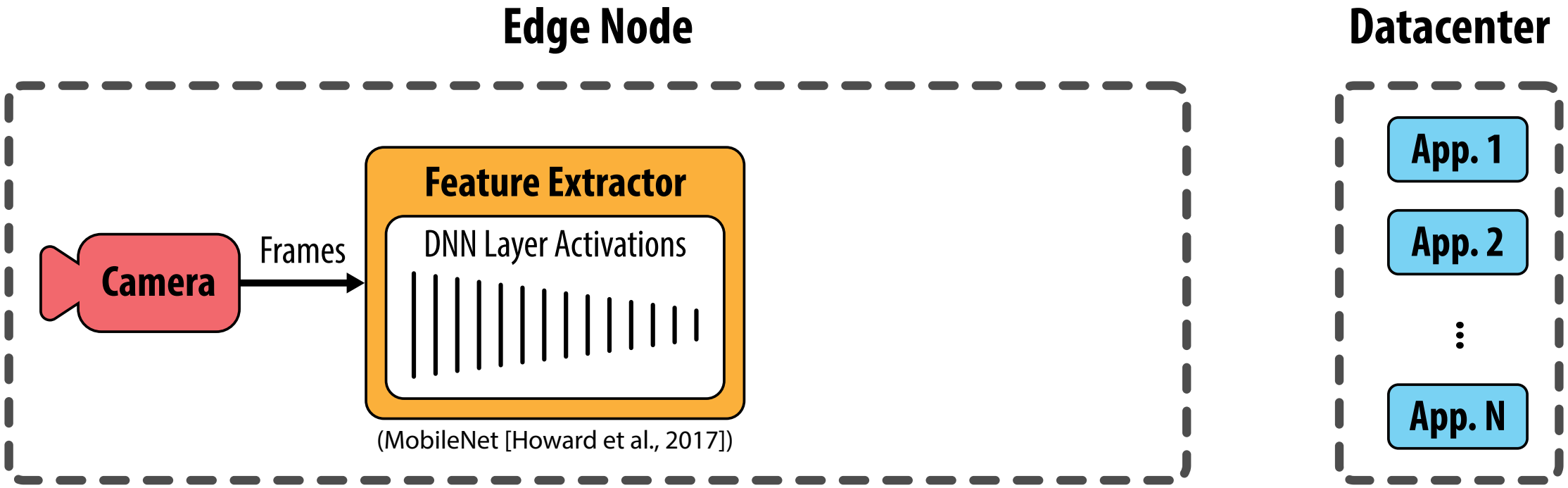
User applications run in the datacenter, filtering happens on the edge

FilterForward at a glance



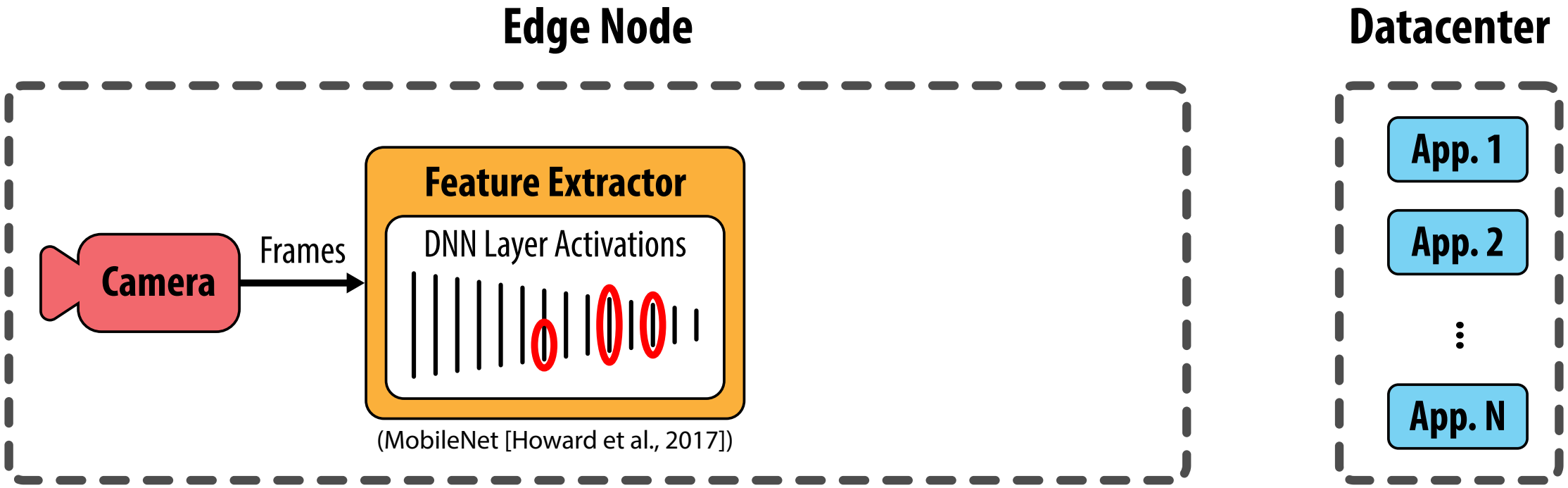
An edge node receives frames from one or more cameras

FilterForward at a glance



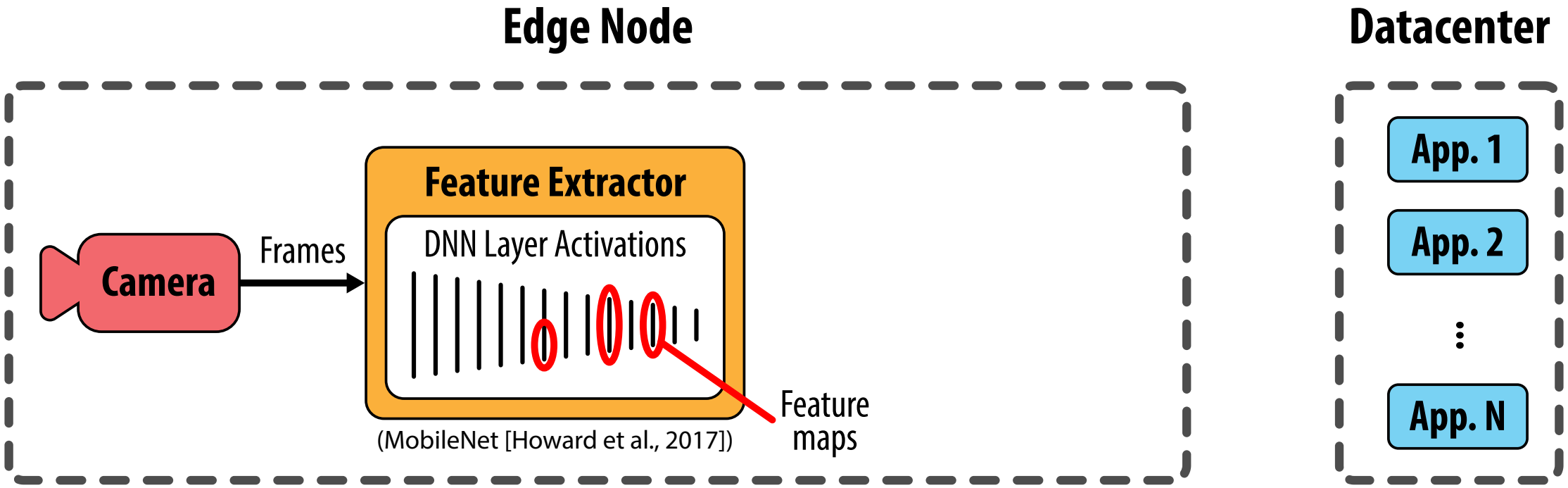
FilterForward amortizes pixel processing by using a ***base DNN*** to extract features

FilterForward at a glance



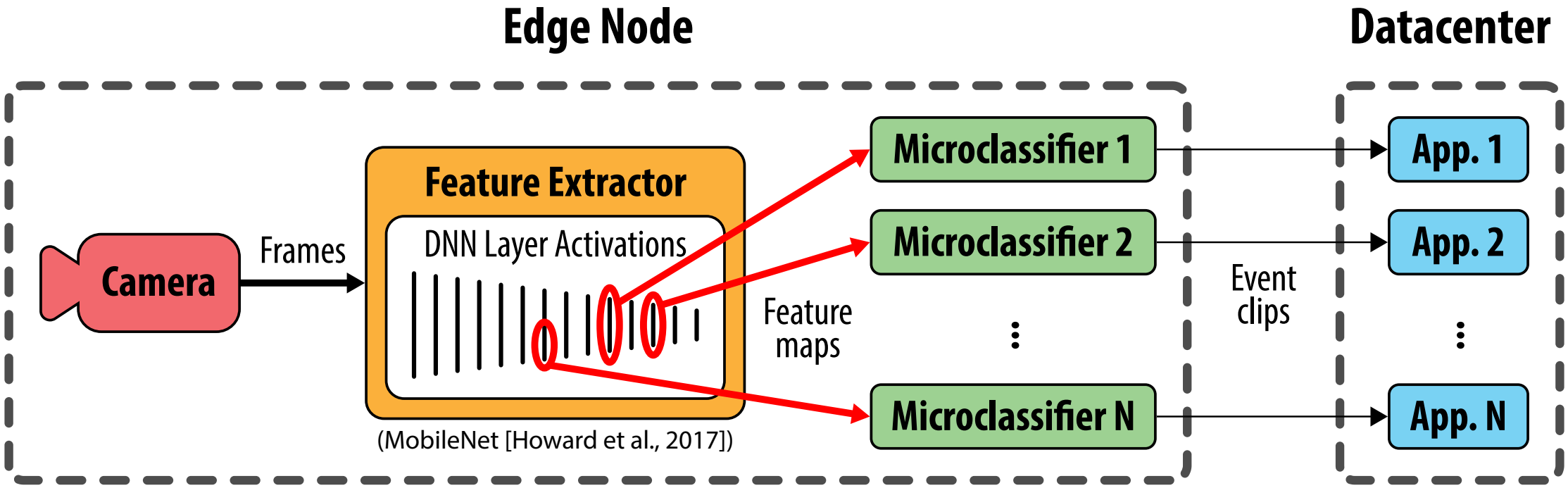
FilterForward amortizes pixel processing by using a ***base DNN*** to extract features

FilterForward at a glance



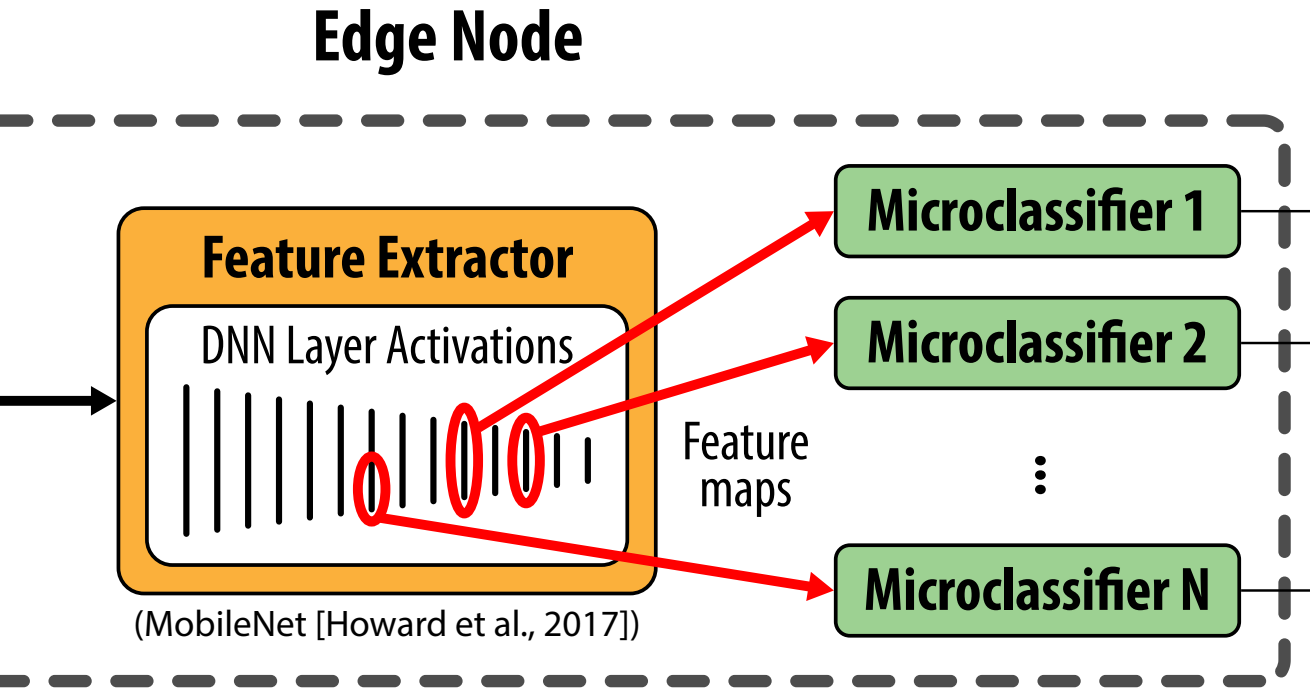
FilterForward amortizes pixel processing by using a ***base DNN*** to extract features

FilterForward at a glance



FilterForward amortizes pixel processing by using a ***base DNN*** to extract features

Microclassifiers are…



Small, individually-trainable neural nets that determine if a frame is relevant

Scalable: ***Shared feature extractor*** amortizes expensive pixel processing

Sensitive to ***fine-grained details*** because they process full-resolution (e.g., 1920x1080) frames and pull from any intermediate layer

Microclassifier architectures

Full-frame object detector (248M mult.-adds)

Feature Extractor → 1x1 Conv → 1x1 Conv → 1x1 Conv → Max → Sigmoid

Localized binary classifier (69M mult.-adds)

Feature Extractor → ConvNet → Sigmoid

Windowed, localized binary classifier (682M mult.-adds)

Feature Extractor → ConvNet → Sigmoid

Evaluation datasets

Two real-world camera feeds:

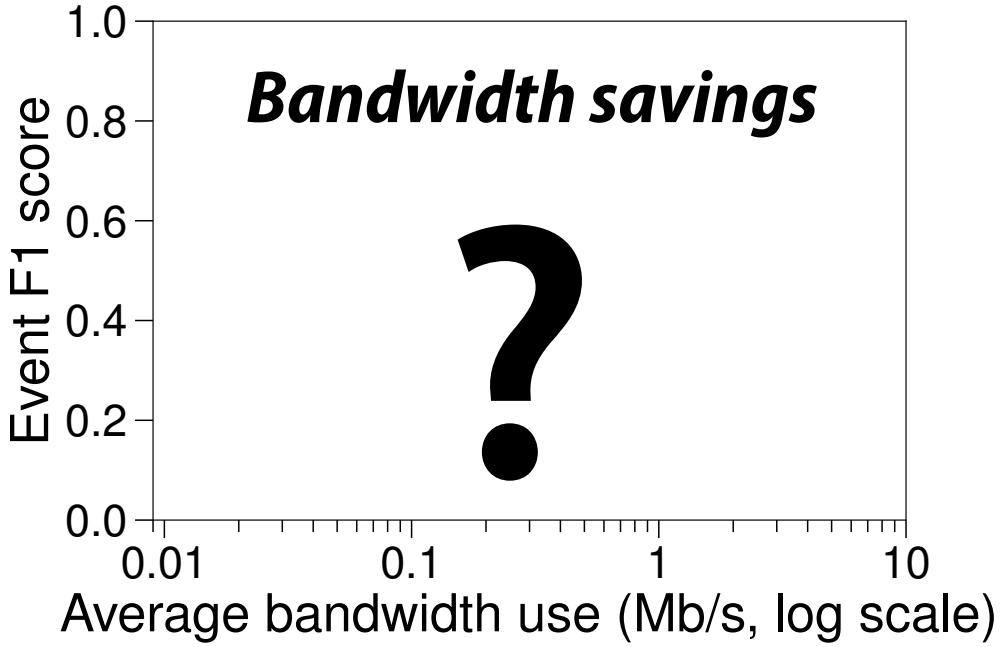
1920x1080

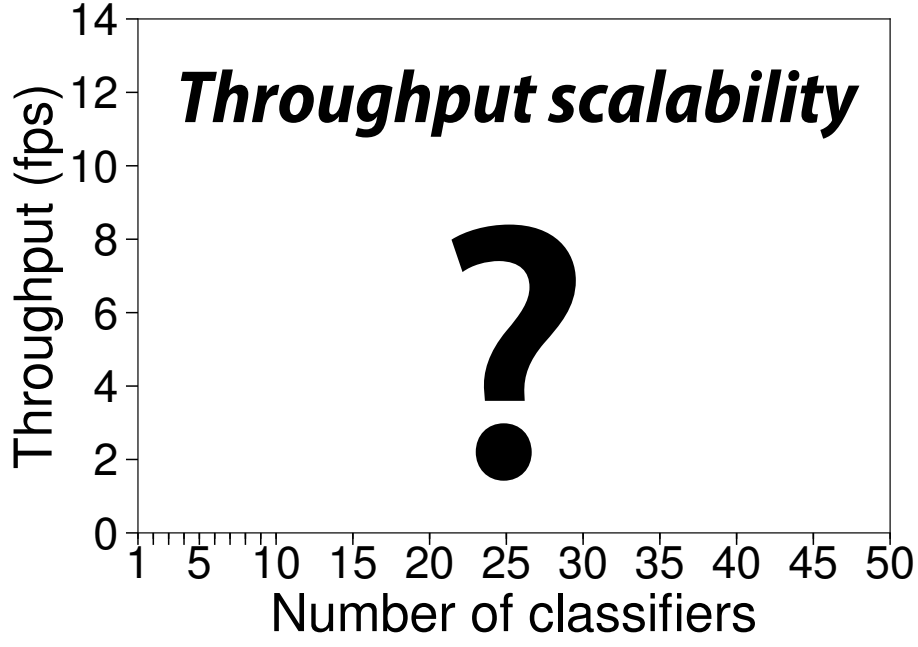
Pedestrians in the crosswalks

2048x850

People wearing red clothes

Evaluation questions



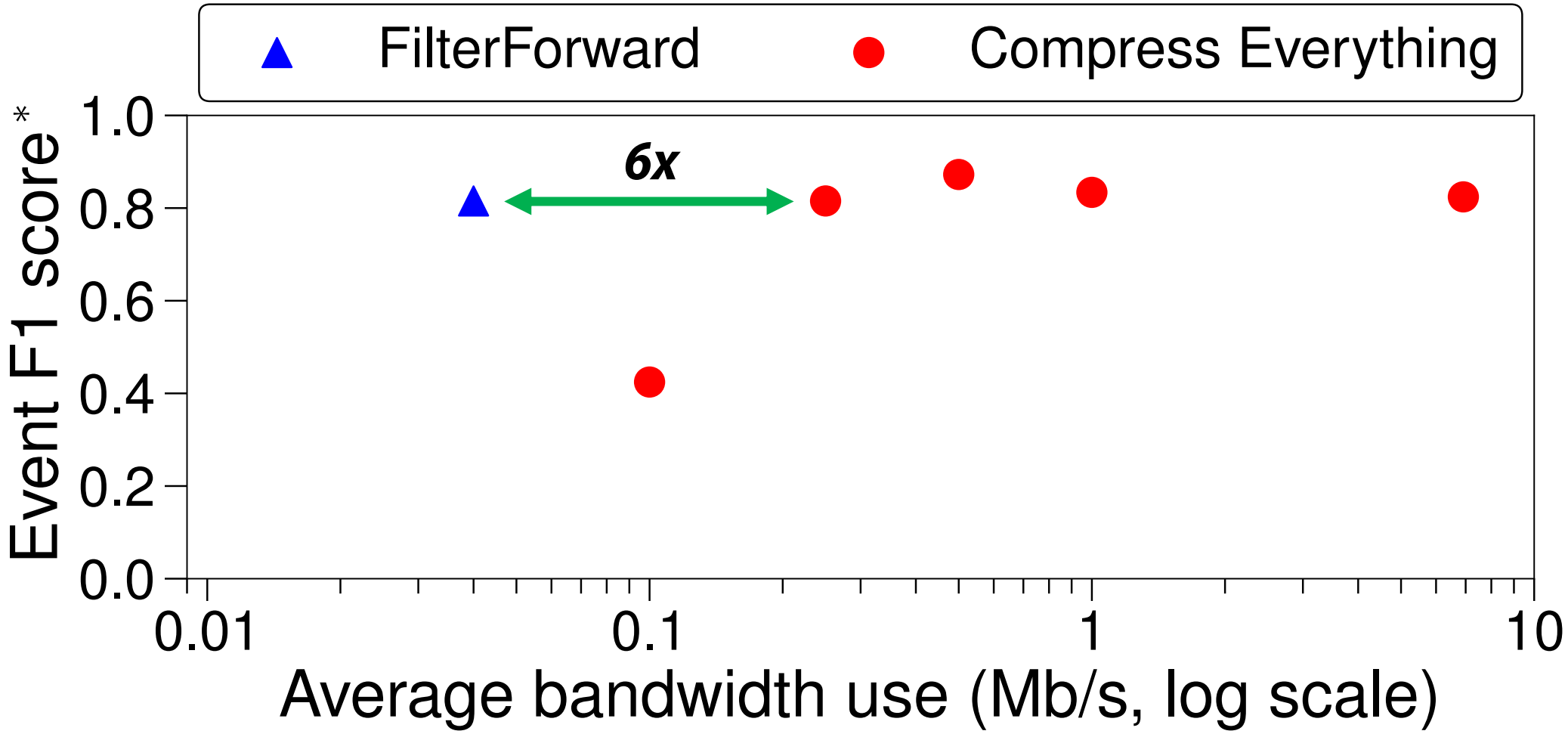


Compare to:

Compress the whole stream using H.264

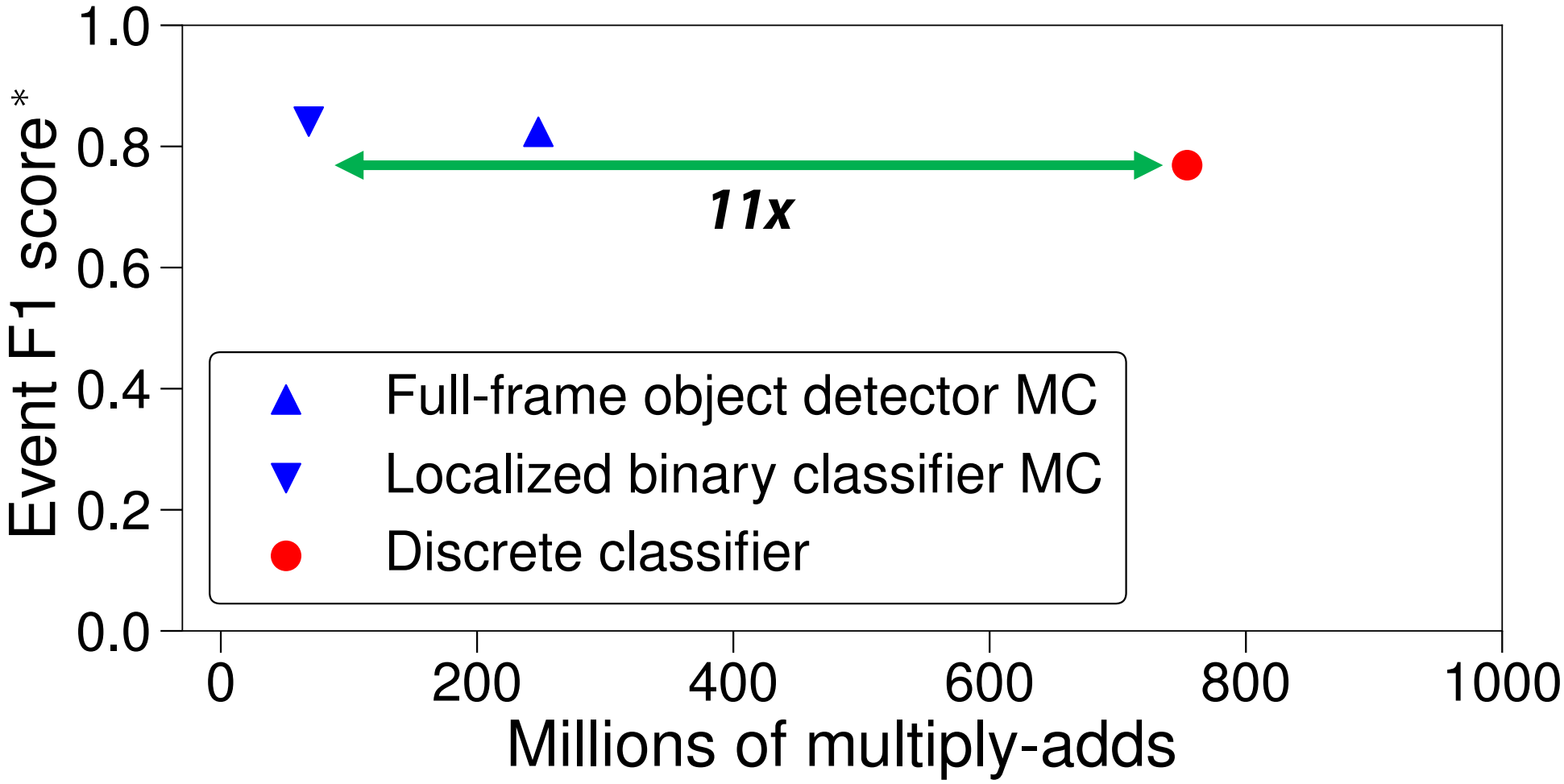
NoScope-style pixel-based classifiers (*discrete classifiers*) [Kang et al., 2017]

FF saves bandwidth without sacrificing accuracy



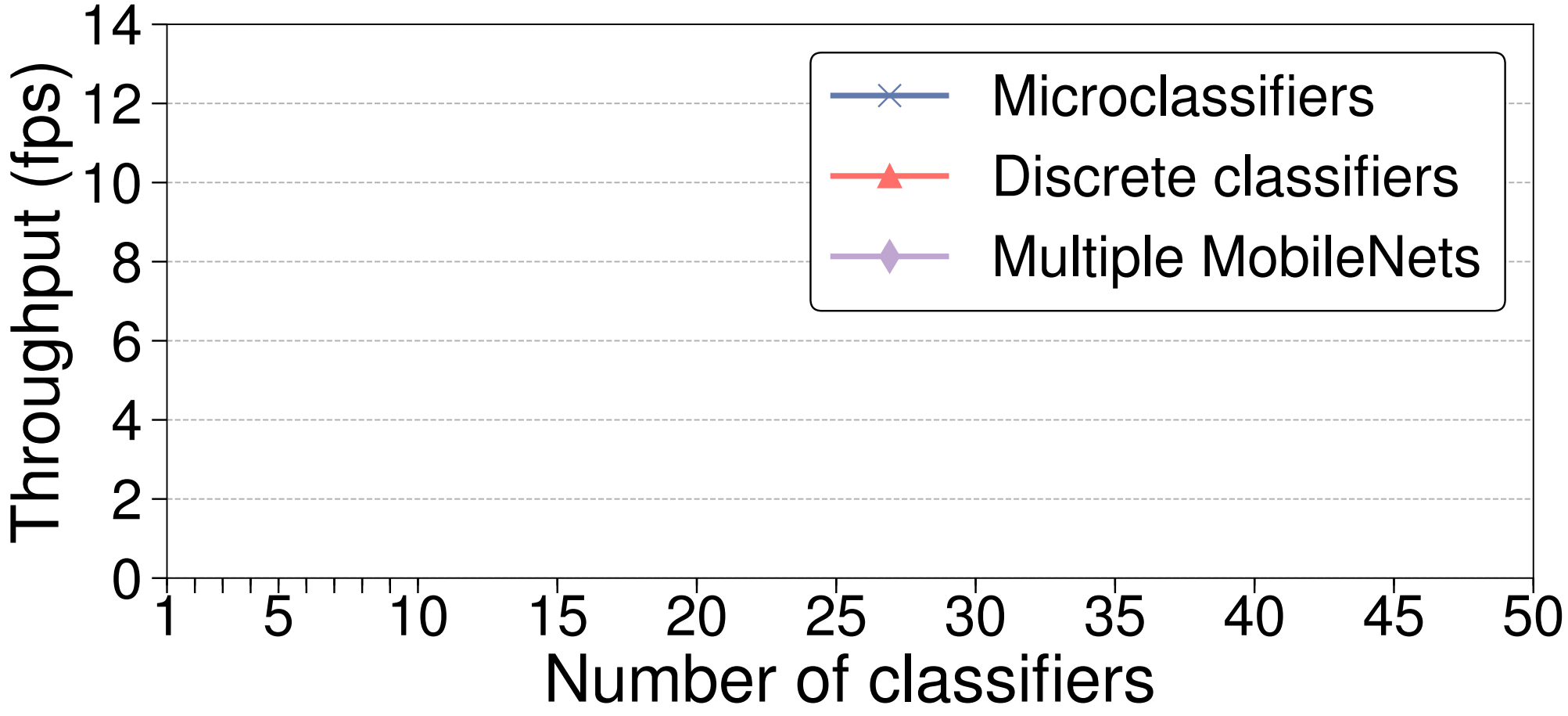
*[Lee et al., SysML 2018]

Microclassifiers are accurate yet lightweight



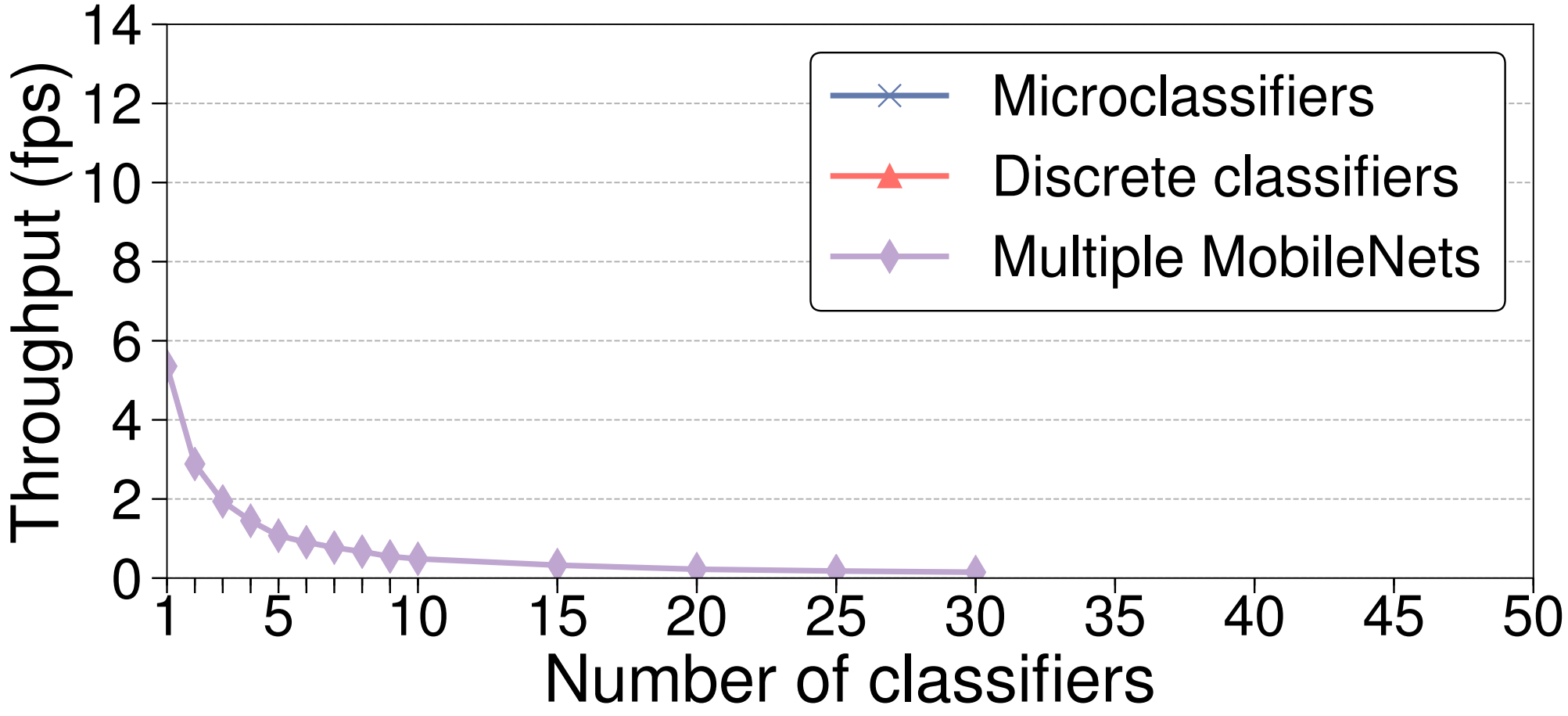
*[Lee et al., SysML 2018]

FilterForward scales to many applications



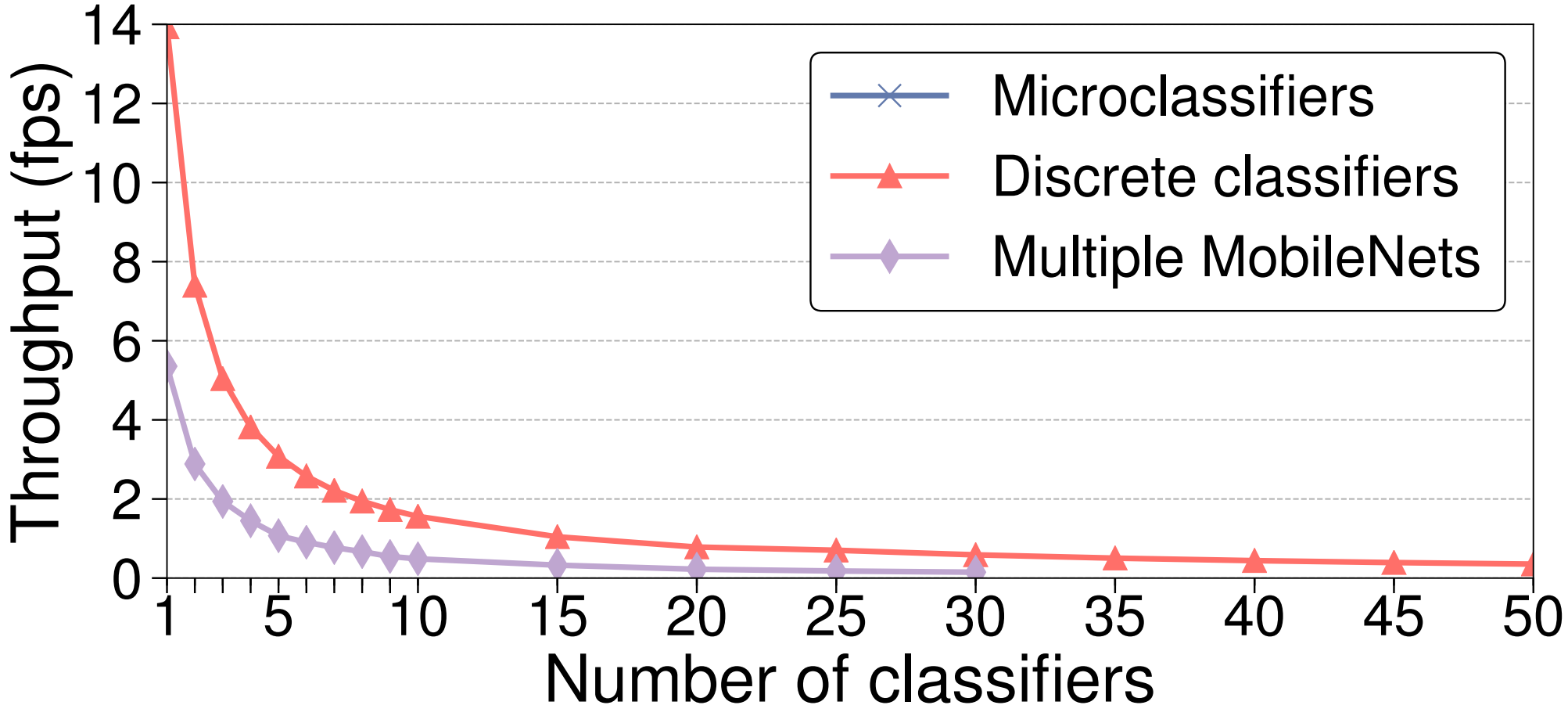
(4-core Intel CPU similar to actual deployments)

FilterForward scales to many applications



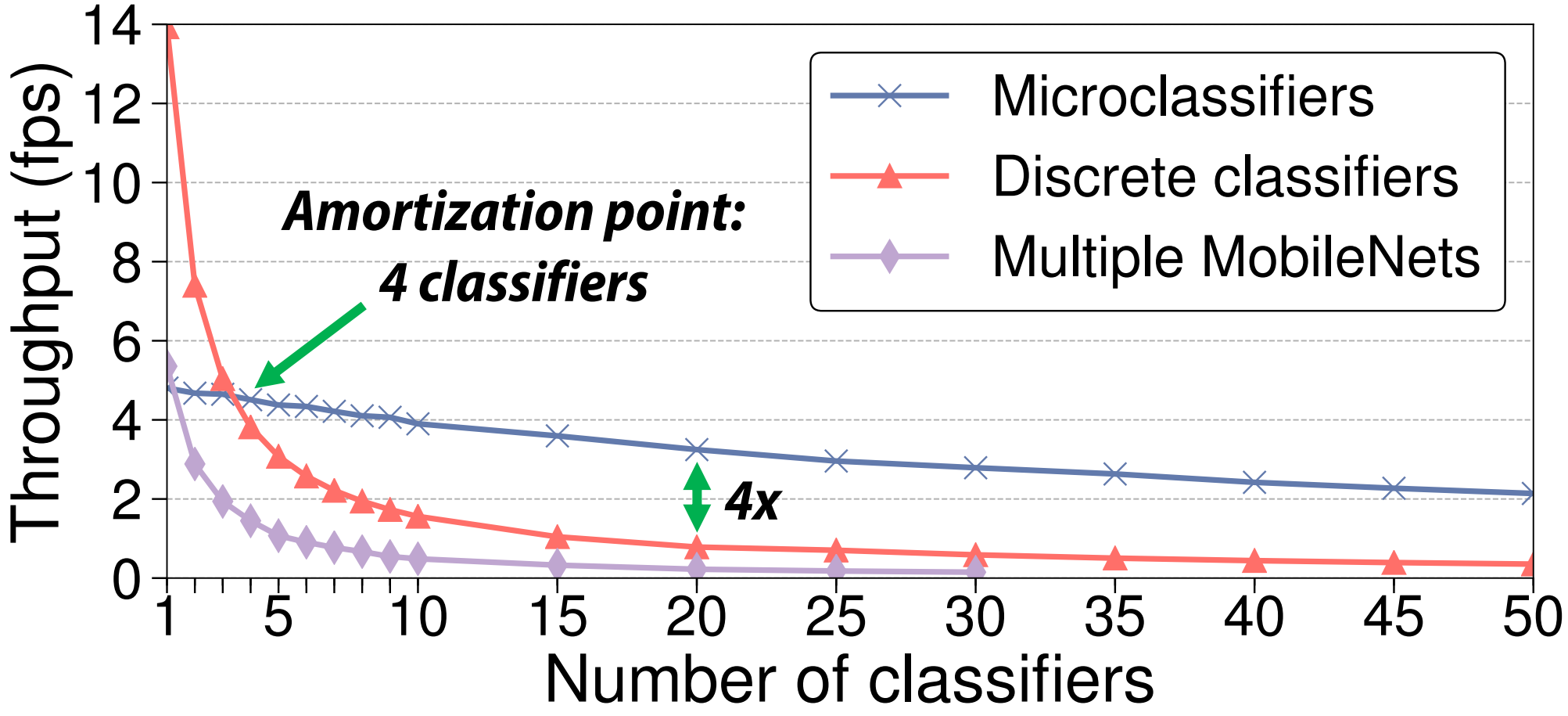
(4-core Intel CPU similar to actual deployments)

FilterForward scales to many applications



(4-core Intel CPU similar to actual deployments)

FilterForward scales to many applications



(4-core Intel CPU similar to actual deployments)

Summary

FilterForward: A new system for efficient video filtering on the edge

- Uses 6x less bandwidth than compression
- With 20 classifiers, 4x higher frame throughput than existing techniques

Microclassifiers: Cheap but accurate application-specific NN filters

- As accurate as pixel-level classifiers with 11x lower marginal cost

Visit us at the poster session, and online at github.com/viscloud/filterforward

Thank you!