Shallow-Deep Networks: Understanding and Mitigating Network Overthinking

Yiğitcan Kaya, Sanghyun Hong, Tudor Dumitraș

University of Maryland, College Park

ICML 2019 - Long Beach, CA

MARYLAND CYBERSECURITY CENTER

What is overthinking?

We, *especially grad students*, often think more than needed to solve a problem.

What is overthinking?

We, *especially grad students*, often think more than needed to solve a problem.

i. Wastes our valuable energy (**wasteful**)

What is overthinking?

We, *especially grad students*, often think more than needed to solve a problem.

i. Wastes our valuable energy (**wasteful**)
ii. Causes us to make mistakes (**destructive**)

Do deep neural networks overthink too?

Without requiring the full depth, DNNs can correctly classify the majority of samples.

Experiments on four recent CNNs and three common image classification tasks

Do deep neural networks overthink too?

Without requiring the full depth, DNNs can correctly classify the majority of samples.

i. Wastes computation for up to **95%** of the samples (**wasteful**)

Do deep neural networks overthink too?

Without requiring the full depth, DNNs can correctly classify the majority of samples.

i. Wastes computation for up to **95%** of the samples (**wasteful**)
ii. Occurs in **~50%** of all misclassifications (**destructive**)

How do we detect overthinking?

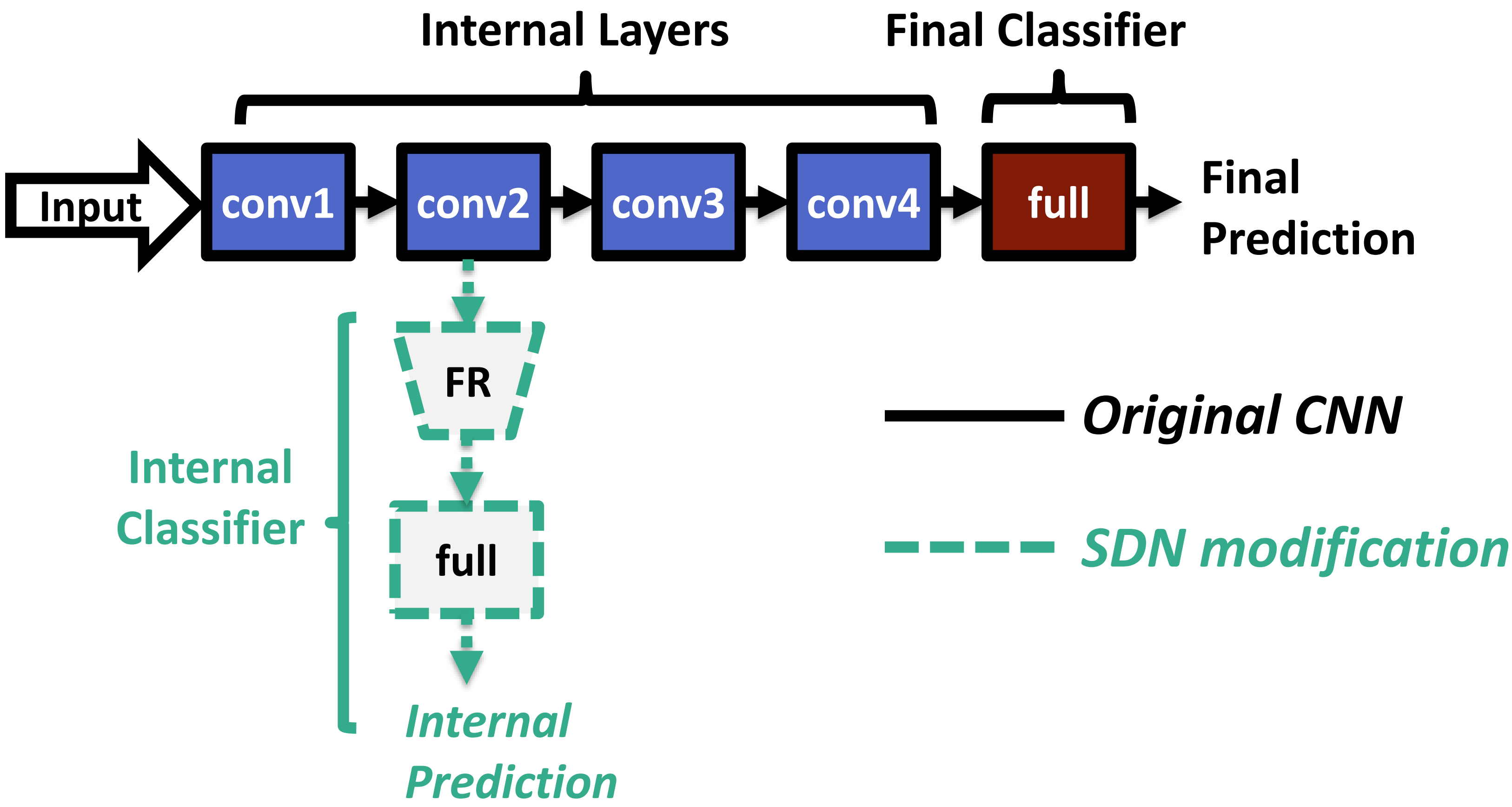
Internal classifiers allow us to observe whether the DNN correctly classifies the sample at an earlier layer.

How do we detect overthinking?

Internal classifiers allow us to observe whether the DNN correctly classifies the sample at an earlier layer.

- Our generic **Shallow-Deep Network (SDN)** modification introduces internal classifiers to *DNNs.*

The SDN modification



Applied to *VGG, ResNet, WideResNet* and *MobileNet.*

The SDN modification

Challenge

How to train accurate internal classifiers?

The SDN modification

Challenge

How to train accurate internal classifiers?

Prior Work

Claims this hurts the accuracy in *off-the-shelf* DNNs

Proposes a *unique* architecture[1]

[1] Huang, Gao, et al. "Multi-scale dense convolutional networks for efficient prediction." *ICLR 2018*

The SDN modification

Challenge

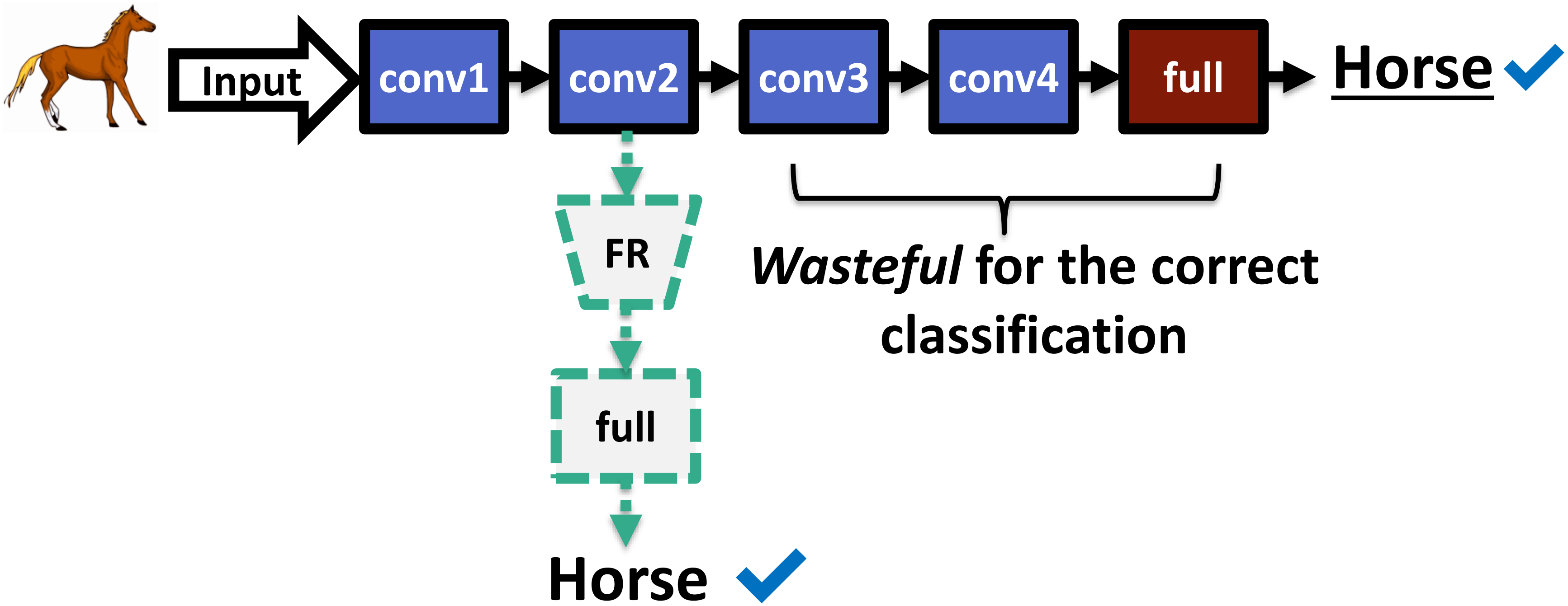
How to train accurate internal classifiers?

Results

Our modification often *improves* the original accuracy by up to **10%**.

(**See our poster**)

The wasteful effect of overthinking



The wasteful effect of overthinking

Challenge

How can we know where in the DNN to stop?

The wasteful effect of overthinking

Challenge

How can we know where in the DNN to stop?

Our Solution

Classification confidence of the internal classifiers

The wasteful effect of overthinking

Our Solution

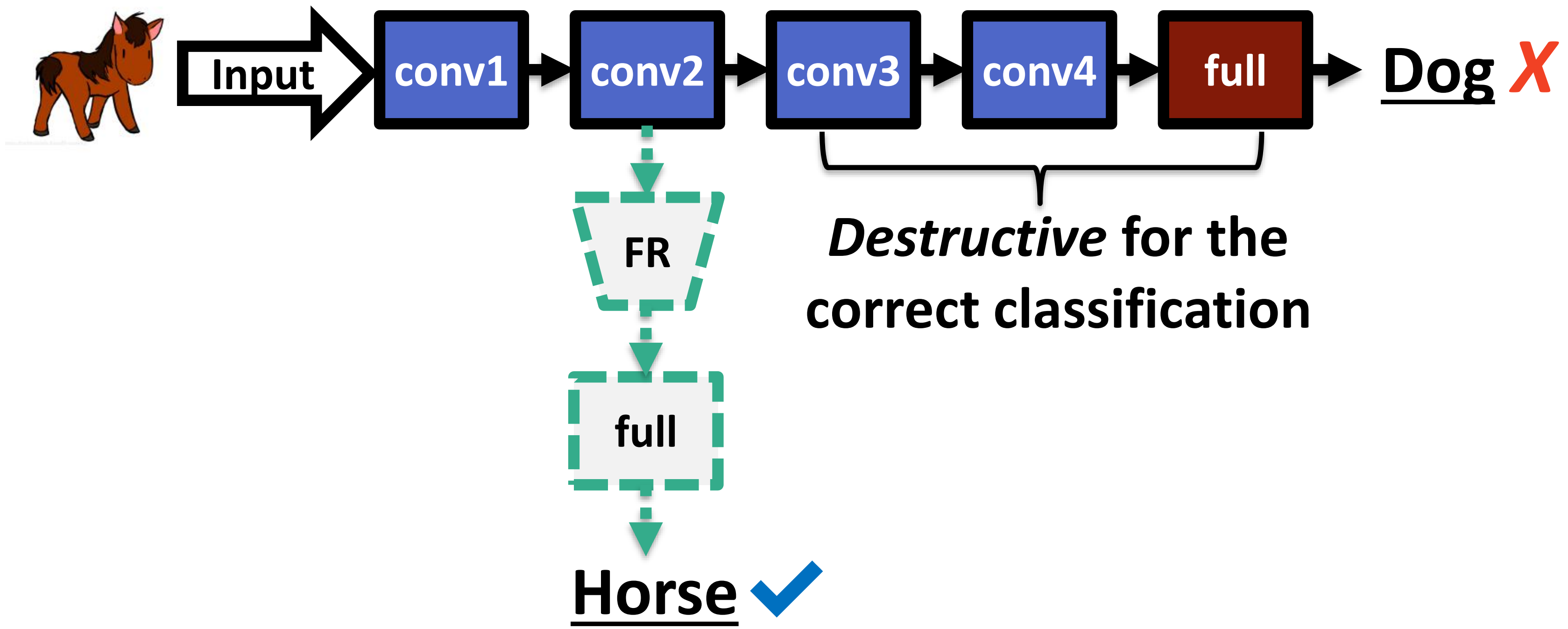
Classification confidence of the internal classifiers

Results

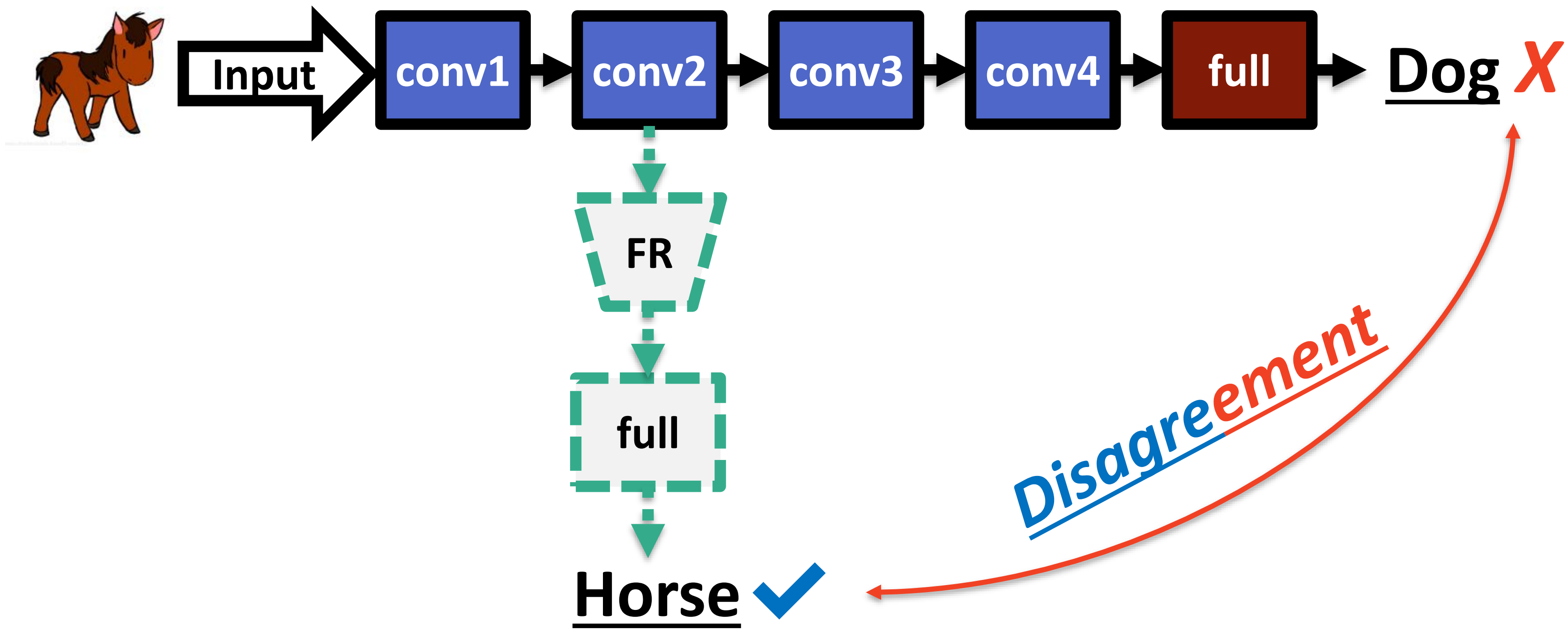
A *confidence-based early exit scheme* reduces the average inference cost by up to **50%**.

(**See our poster**)

The destructive effect of overthinking



The destructive effect causes disagreement



The destructive effect causes disagreement

Challenge

How can we quantify the internal disagreement?

Our Solution

The *confusion* metric

The destructive effect causes disagreement

<u>Our Solution</u>

The confusion metric

<u>Results</u>

Confusion indicates *whether a misclassification is likely.*

Confusion is a *reliable error indicator.*

(**See our poster**)

The destructive effect causes disagreement

Our Solution

The confusion metric

Results

Backdoor attacks [2] also increase the confusion of the victim DNN for malicious samples.

(**See our poster**)

[2] Gu, Tianyu, et al. "BadNets: Evaluating Backdooring Attacks on Deep Neural Networks." *IEEE Access* 7 (2019): 47230-47244.

Implications

- Eliminating overthinking would lead to *a significant boost* in accuracy and inference-time.

- We need DNNs that can *adjust their complexity* based on the required feature complexity.

For more details, visit our website

http://shallowdeep.network

Thank you!

Don't overthink! Come and see our poster!

Pacific Ballroom – Poster #24 – 06:30-09:00 PM