Efficient Learning of CNNs using Patch Based Features

A theoretical perspective on patch-based models

Alon Brutzkus [1] Amir Globerson [1] Eran Malach [2]
Alon Regev Netser [2] Shai Shalev-Shwartz [2]

[1]Tel Aviv University, Israel
[2]The Hebrew University of Jerusalem, Israel

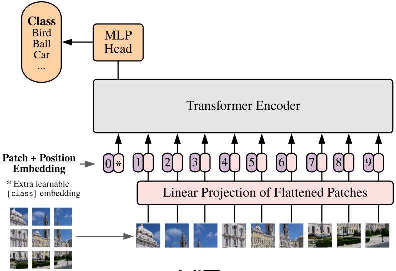
June 27, 2022

Motivation: Understanding patch-based models

- Recently, learning with patch-based representations has become increasingly popular for solving visual tasks.

Motivation: Understanding patch-based models

- Recently, learning with patch-based representations has become increasingly popular for solving visual tasks.



ViT

Motivation: Understanding patch-based models

Recently, learning with patch-based representations has become increasingly popular for solving visual tasks.

ViT

MLPMixer

Motivation: Understanding patch-based models

Recently, learning with patch-based representations has become increasingly popular for solving visual tasks.

ViT

MLPMixer

Our goal is to gain a deeper understanding of these models, by studying a simple patch-based model from a theoretical perspective

The Model[1]

[1]Originally proposed by Coates et al. (2011) and improved by Thiry et al. (2021).

The Model[1]

Unsupervised Stage:
Obtain a patches dictionary D by clustering patches from the data.

[1]Originally proposed by Coates et al. (2011) and improved by Thiry et al. (2021).

The Model[1]

Unsupervised Stage:
Obtain a *patches dictionary D* by clustering patches from the data.

Patch-Based Image Embedding:
Define a patch-embedding ${}_D(\ )$ mapping patches to R^t, which induces an image-embedding $\Phi_D(\ )$.

[1] Originally proposed by Coates et al. (2011) and improved by Thiry et al. (2021).

The Model[1]

Unsupervised Stage:
Obtain a *patches dictionary D* by clustering patches from the data.

Patch-Based Image Embedding:
Define a patch-embedding $_{D}(\ )$ mapping patches to R^{t}, which induces an image-embedding $\Phi_{D}(\ )$.

Supervised Stage:
Train a linear model on top of the fixed embedding on labeled data.

[1]Originally proposed by Coates et al. (2011) and improved by Thiry et al. (2021).

Provable Efficient Learning Under Some Assumptions

We prove the algorithm learn efficiently assuming the patches distribution has a low covering-number (e.g., low intrinsic dimension).

Provable E cient Learning Under Some Assumptions

We prove the algorithm learn e ciently assuming the patches distribution has a low covering-number(e.g., low intrinsic dimension).

[redacted]

Fix ; 2 $(0;1)$. Let I be a distribution realizable by a shallow CNN. Le N_0 be the covering-number of the patches-distribution. Then, running th algorithm with a dictionary of size N_0 and poly$(1= ; 1= ; N_0)$ samples returns w.p. at least1 a hypothesish s.t. $\Pr_{(x;y)\ \mathrm{I}}\ [h(x) \ne y]$.

Provable E cient Learning Under Some Assumptions

- We prove the algorithm learn e ciently assuming the patches distribution has a low covering-number(e.g., low intrinsic dimension).

[redacted]

Fix ; 2 (0; 1). Let I be a distribution realizable by a shallow CNN. Le N_0 be the covering-number of the patches-distribution. Then, running th algorithm with a dictionary of size N_0 and poly(1= ; 1= ; N_0) samples returns w.p. at least 1 a hypothesis h s.t. $\Pr_{(x;y)\ I}$ [h(x) 6 y] .

- We also suggest a new embedding and prove it provides e cient learning under some assumption on the target function.

Performance Analysis

The algorithm equipped with our $\text{embedding}_{\text{full}}$ outperforms the previously proposed $\text{embedding}_{\text{hard}}$, and even a vanilla CNN.

Performance Analysis

- The algorithm equipped with our embedding Φ_{full} outperforms the previously proposed embedding Φ_{hard}, and even a vanilla CNN.

- Running the algorithm with Φ_{full} in a layer-by-layer fashion results in a deep model which gives further improvements, while the original algorithm (with Φ_{hard}) does not scale with depth.

Verifying Distributional Assumption

- We verify that the distributional assumptions hold on real world data by experimenting on CIFAR-10 and ImageNet datasets.

Verifying Distributional Assumption

- We verify that the distributional assumptions hold on real world data by experimenting on CIFAR-10 and ImageNet datasets.
- We observed that patches sampled from the data are clustered together.

Thank you!

- We provide new understanding of patch-based representations.
- If you find our work interesting please visit our poster.[2]

[2] You can also email us at alon.netser@mail.huji.ac.il or eran.malach@mail.huji.ac.il