

2024/12/03

- koizumi

Focus Your Distribution: Coarse-to-Fine Non-Contrastive Learning for Anomaly Detection and Localization

<https://arxiv.org/abs/2110.04538>

(ICME 2020)

異常検知タスク

- Coarse Alignment

入力画像内のオブジェクトの向き・位置を揃える (Affine変換)

特徴量抽出 (ResNetパラメタ固定) + 位置調整 (Affine変換)

各正常画像の特徴量の分布が小さくなるように学習

- Fine Alignment

特徴量のコピーとパッチごとにシャッフルしたものに分ける

上記二つのSimilarityを取ることでテスト時に異常を判定

片方の勾配を止めることで学習崩壊を防ぐ

- Loss

End-to-Endで最適化

$$\mathcal{L}_{total}(\cdot; \theta_{h,f,g}, \mathcal{T}_{\theta}) = \lambda_1 \cdot \mathcal{L}_{ICA} + \lambda_2 \cdot \mathcal{L}_{FAS}$$

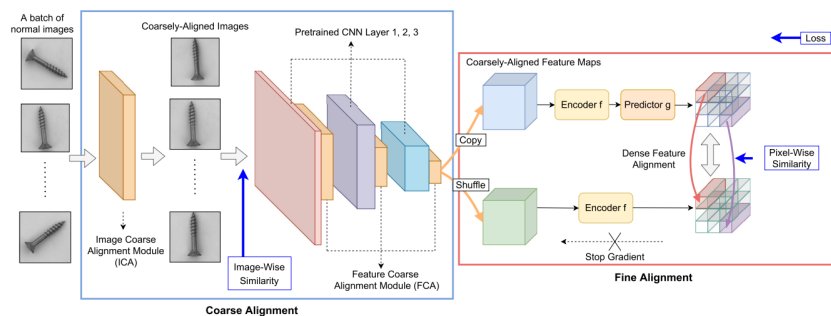


Figure 1: The whole architecture of the proposed approach. The coarse alignment module standardizes the image-wise and feature-wise correspondence among pixels, while the fine alignment module densely maximizes the similarity of these pixels in a batch. We train these two modules with non-contrastive learning method in an end-to-end manner.

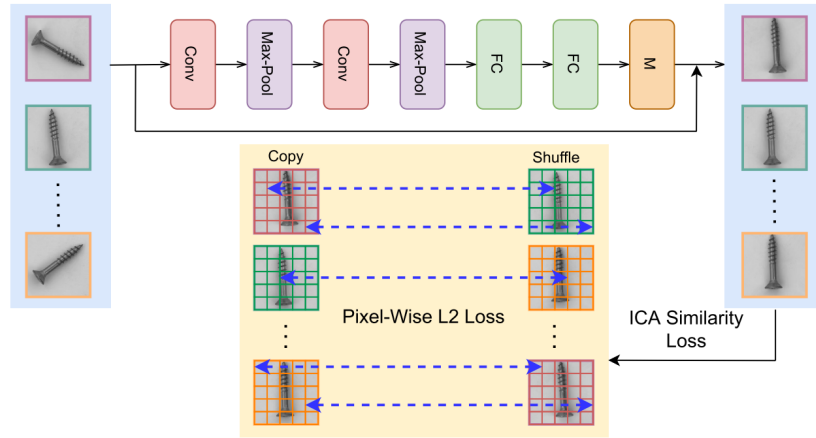


Figure 4: The architecture of Image-level Coarse Alignment module, ICA, supervised by similarity loss, learns the rotation transformation parameters from input images and then applies the affine matrix to them to get batch images in same pose.

- Jiu

LoRA: Low-Rank Adaptation of Large Language Models

<https://arxiv.org/abs/2106.09685>

(ICLR 2022)

大規模言語モデルに対する追加学習手法

- 大規模モデルのfine-tuningはコストが高い
- 低ランク行列を用いて更新パラメータ差分を学習

fine-tuningで必要となるベクトルの次元は元のモデルの次元と比較して少なく済む
(<https://arxiv.org/abs/2012.13255>)

$$h = W_0 x + \Delta W_x = W_0 x + B A x$$

W_0 は固定、 A, B を学習、 A はガウス初期化、 B はゼロ初期化

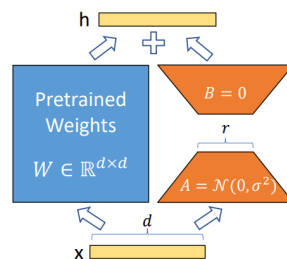


Figure 1: Our reparametrization. We only train A and B .

- fujinami

D4: Detection of Adversarial Diffusion Deepfakes Using Disjoint Ensembles

<https://arxiv.org/abs/2202.05687>

(WACV2024)

Adversarial Deep Detection

- non-adversarialなディープフェイク検出器は知覚できない敵対的な摂動によって予測を操作される
- 従来のAdversarialなディープフェイク検出器は大きく学習時間と推論時間の二つの防御策を取る
training time：学習時に敵対的な例を用いる / アンサンブルを用いる → 堅牢度の大幅な向上には至らない
Inference time：敵対的な摂動の影響を除去するほうに重点を置く → 計算量が多く、時間がかかる
- 複数のモデルで分割された異なる周波数帯域を用いることでディープフェイクの堅牢度を向上
周波数空間の冗長性により分割が可能
全てのモデルが等しい堅牢度となるよう顕著性 s_i を計算

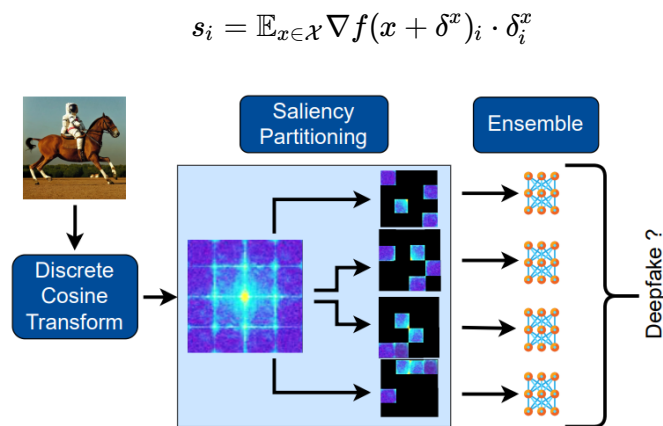


Figure 3. The processing pipeline of D4. It partitions the DCT spectrum of an image into disjoint partitions using a saliency-based approach. Each frequency partition is fed to a separate model that is adversarially trained. A voting mechanism over the ensemble decides the output.

- tanaka

Enhanced Quantified Local Implicit Neural Representation for Image Compression

<https://ieeexplore.ieee.org/document/10323534>

(IEEE Signal Processing Letters vol. 30)

Implicit Neural Representation for Image

- 従来手法：COOL-CHIC
隣接ピクセル間の関係を活用できていない
量子化誤差による影響
- LEM
MLPの後に隣接ピクセル間の関係を利用するLEM (CNN) を導入

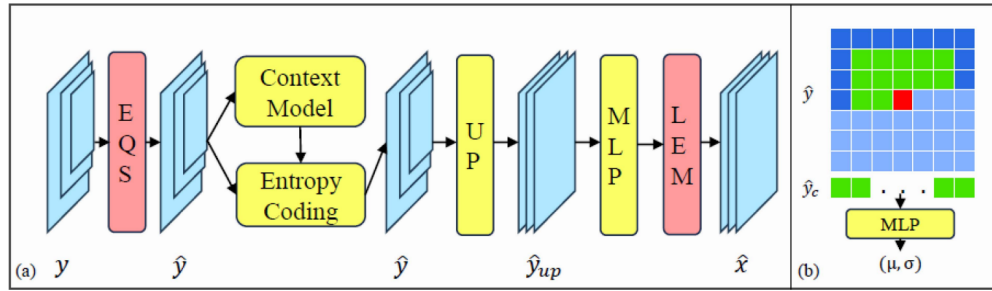


Fig. 1. Part (a) is the framework of EQLINR. y is a learnable latent representation. EQS is the enhanced quantization scheme. \hat{y} is the quantified y . UP means upsample. \hat{y}_{up} is the upsampled \hat{y} . MLP means multilayer perceptrons. LEM is the local enhancement module. \hat{x} is the reconstructed image from LEM. Part (b) is the autoregressive context model. \hat{y} is the input latent representation. \hat{y}_c is the context. (μ, σ) is the output parameters for Gaussian distribution.

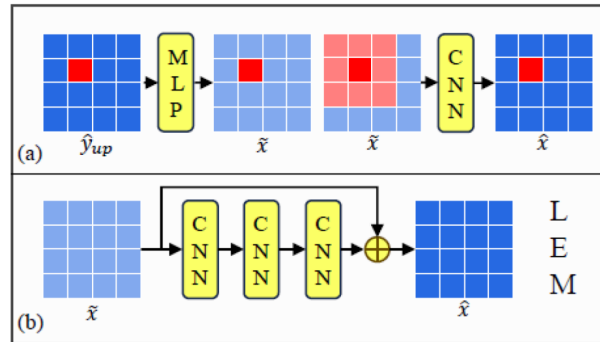


Fig. 2. Different utilization of local relationships by CNN and MLP.

- EQS

COOL-CHIC：一様ノイズを使用 → 量子化誤差との差がまだ大きい

提案：一様ノイズを加え、stochastic gumbel annealing(SGA)とdynamic temperature regularizationにより量子化誤差とのギャップを埋める