Image Attribution is the process of detecting if an image is synthetic, and if yes, which synthesizer generates that image.

Challenges:
- Attribution to model architecture is more challenging than to model instance → require generalization to unseen training categories.
- Robustness to noises: perturbations during online redistribution typically distort high frequency details which would be useful for image attribution.

Motivation:
- Existing attribution datasets [5] are non-deterministic, lack of semantic diversity, source model’s quality and levels of perturbations.
- The needs of a new benchmark for model architecture attribution instead of model instance attribution.

The Attribution88 dataset:
- 1,056,000 RGB images of size 128x128.
- 8 source classes (real + 7 GAN architectures) x 11 semantics.
- 77 GAN models have been collected or trained using LSUN real images.
- Data cleaning: removing artifacts and balancing diversity.
- Noises: 19 ImageNet-C random perturbation sources.
- Deterministic test set: semantics (6 seen + 5 unseen), perturbations (15 seen + 4 unseen).

RepMix Architecture:
\[ f(x_i, x_j, \alpha) = f_t(\alpha f_b(x_i) + (1 - \alpha) f_b(x_j)) \]

where \( f_t \) and \( f_b \) are top and bottom layers; \( \alpha \) is drawn randomly from Beta distribution; \( x_i \) and \( x_j \) are 2 input images.

- Mix random image features of different semantics and sources (GANs or real).
- Predict the source mixture ratio.
- Mixing at top fully connected layer gives the best performance.

Compound loss:
- Account for the hierarchical nature of attribution: image → real or fake (synthesised) → generator sources,
- Consist of real/fake detection loss and attribution loss,
- Detection (real/fake) scores factor in attribution scores,
- Weighted cross-entropy losses.

5. References


5. References


5. References