On the Fly Neural Style Smoothing for Risk-Averse Domain Generalization

Akshay Mehra[1], Yunbei Zhang[1], Bhavya Kailkhura[2], and Jihun Hamm[1]
[1]Tulane University, [2]Lawrence Livermore National Laboratory

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

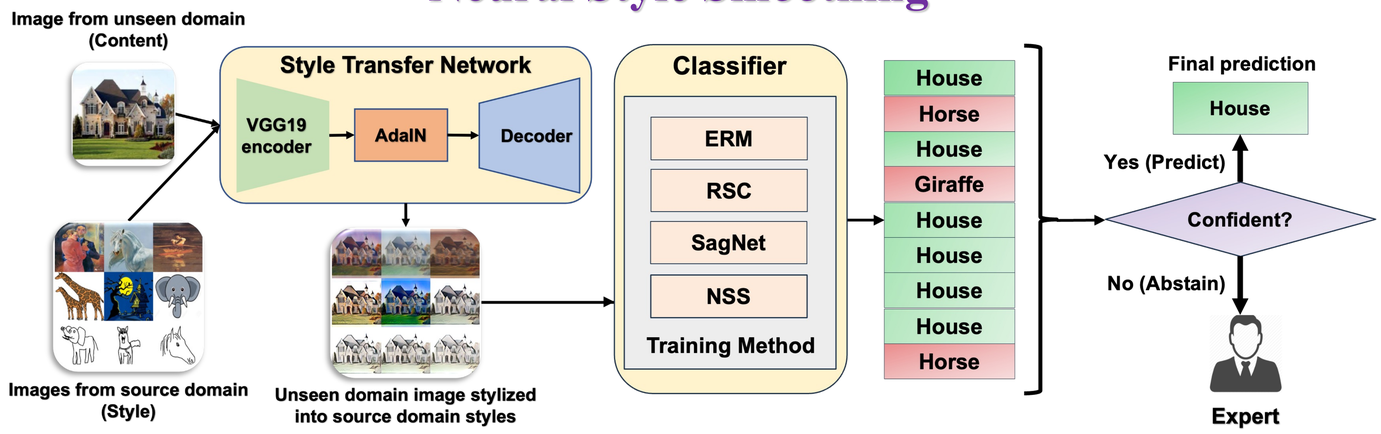
Motivation

- Machine learning (ML) models suffer a significant accuracy degradation on data from unseen domains.
- Recent works demonstrated the use of domain-specific information (e.g., image styles) for predictions as a potential cause for this accuracy degradation.
- This makes ML models unreliable for deployment in high-risk scenarios where misclassifications can have catastrophic consequences.

Contributions

- We propose an efficient inference procedure to obtain risk-averse predictions on the fly from a model by requiring only black-box access to it.
- We propose a training procedure relying on losses that enforce prediction consistency on random stylization of the training data.
- We demonstrate the effectiveness of our inference and training methods on benchmark datasets and their variations.

Neural Style Smoothing



- Let $\mathcal{X}$, $\mathcal{Y}$ be the data domain/labels, f be the classifier, and x_c/x_s be the content/style images, g/h be the encoder/decoder, then $h(t)$ is the AdaIn image with $t := AI(x_c, x_s) = \sigma(g(x_s))\left(\frac{g(x_c)-\mu(g(x_c))}{\sigma(g(x_c))}\right) + \mu(g(x_s))$.

- Given a classifier $f: \mathcal{R}^d \to \mathcal{Y}$, prediction of a style smoothed classifier $\psi: \mathcal{R}^d \to \mathcal{Y}$ on a sample x is given by

$$\boldsymbol{\psi(x) := \arg\max_{y \in \mathcal{Y}} P\left(f\left(h(t)\right) = y\right)},$$

where $t = AI(x, x_s)$, $x_s \sim P_S$ and P_S is the source distribution.

def $get\_counts(x, f, g, h)$:
  set cls_counts to zeros.
  for $i = 1, \ldots, n$ **do**
    Sample x_s^i from P_S
    $t = AI(x, x_s^i; g)$
    $x_{stylized} = h(t)$
    $pred = f(x_{stylized})$
    $cls\_counts[pred]$ += 1
  return cls_counts

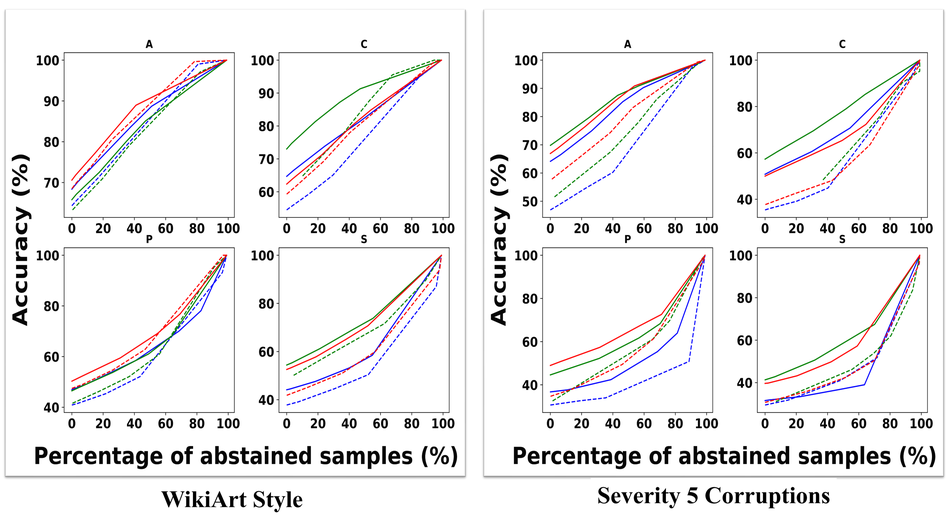
def $TT\_NSS(x, f, g, h, n, \alpha)$:
  $cls\_counts = \boldsymbol{get\_counts}(x, f, g, h)$
  $c_{max} = \arg\max_y cls\_counts$
  $n_{max} = cls\_counts[c_{max}]$
  if $\frac{n_{max}}{n} < \alpha$ ***then***
    return $ABSTAIN$
  else:
    return c_{max}

- NSS-based losses to improve risk-averse predictions.
  - **Style Augmentation**: $\mathbb{E}_{x_s \sim P_S}[\ell(f(h(t)), y)]$.
  - **Style Consistency**: $\mathbb{E}_{x_s \sim P_S}[KL(\bar{F}(x)) || F(h(t))] + H(\bar{F}(x), y)$, where $F: \mathcal{R}^d \to \Delta^{K-1}$ is the soft classifier of f and Δ^{K-1} is the probability simplex in $\mathcal{R}^K$, $H(\cdot)$ is the entropy and $\bar{F}(x) = \mathbb{E}_{x_s \sim P_S}[F(h(t))]$.

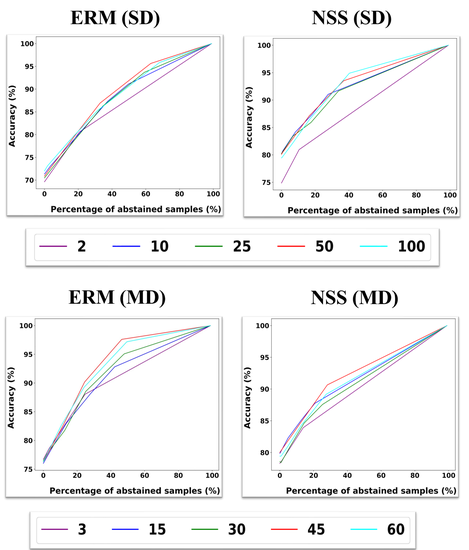
Test-Time Neural Style Smoothing (**TT-NSS**)

Key Results

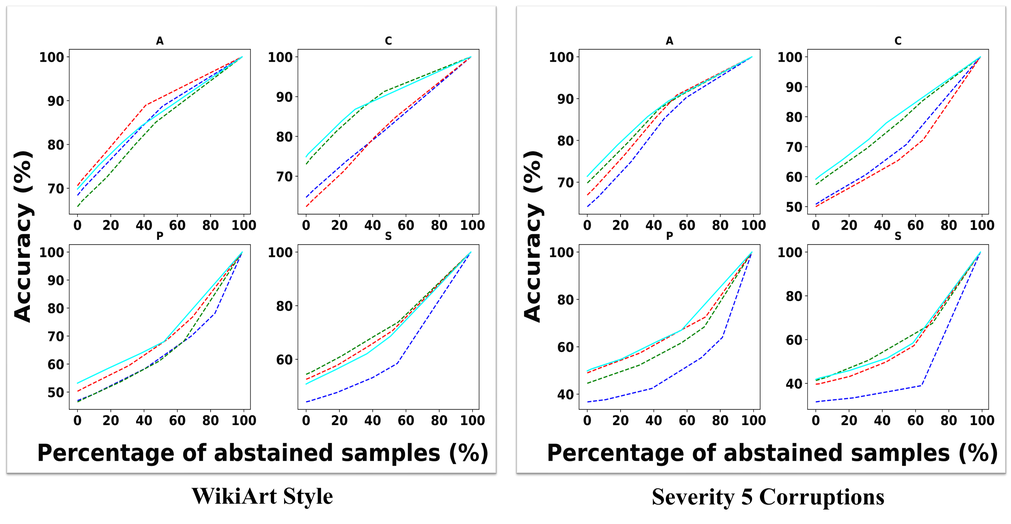
- TT-NSS produces better risk-averse predictions than the confidence-based abstaining mechanism.



- Effect of the number of re-stylizations, n, in TT-NSS.



- NSS trained models achieve better risk averse predictions than models trained with ERM.



This work was supported in part by the NSF EPSCoR-Louisiana Materials Design Alliance (LAMDA) program #OIA1946231 and in part by the Harold L. and Heather E. Jurist Center of Excellence for Artificial Intelligence at Tulane University and was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344 and was supported by the LLNL-LDRD Program under Project No. 23-ERD-030.