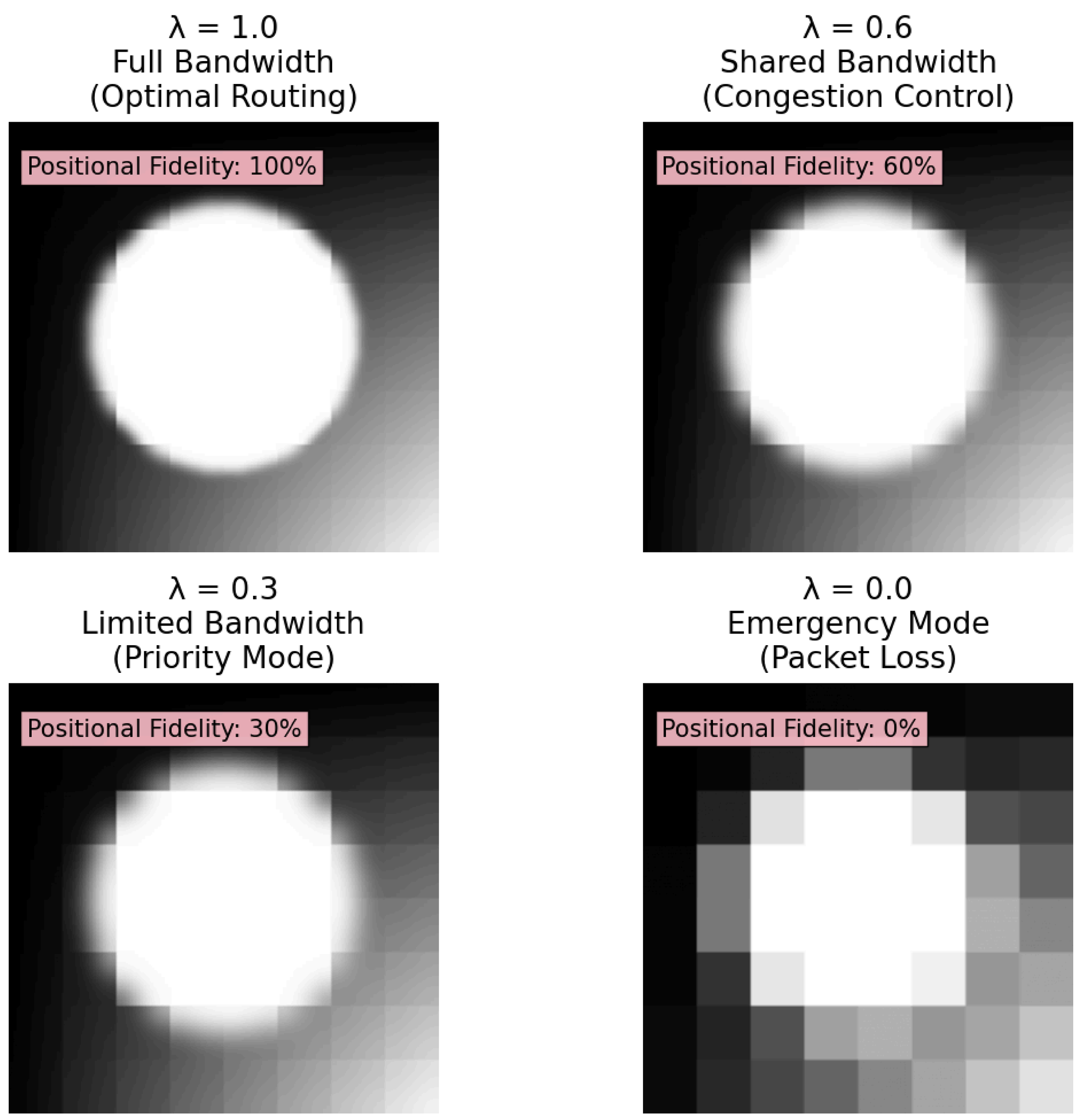
Adaptive Positional Encoding with Regularization for Robust Edge Computing Applications

Chiara Camerota, Flavio Esposito

Saint Louis University, USA

Intuition

This research presents a modified Visual Transformer (ViT) that utilizes Rotational Positional Encoding (RoPE), which considers network and device metadata (λ).



If $\lambda = 1$, the spatial relationships are identical to classical RoPE; if the metadata indicates a bad network and device condition, the spatial structure is simplified.

Methodology

We propose a regularization of the angle of RoPE:

$$\tilde{\theta} = \theta\tilde{\lambda}, \quad \tilde{\lambda} \in [0,1] \rightarrow \boldsymbol{R}^{\text{Rope}}(\tilde{\theta})$$

Global adjustment

$$\tilde{\lambda} = \phi\gamma(\xi) + (1-\phi)\Gamma, \quad \phi \in [0,1]$$

Network quality $\gamma(\xi) \in [0,1]$

Local health $\text{Sigmoid}(\sum w_p \alpha_p)$ where α_p represents battery, CPU, and GPU status

Assuming $\mathcal{N}(R^d_{\Theta,\text{true}}, \sigma^2 I)$ the optimal value is:

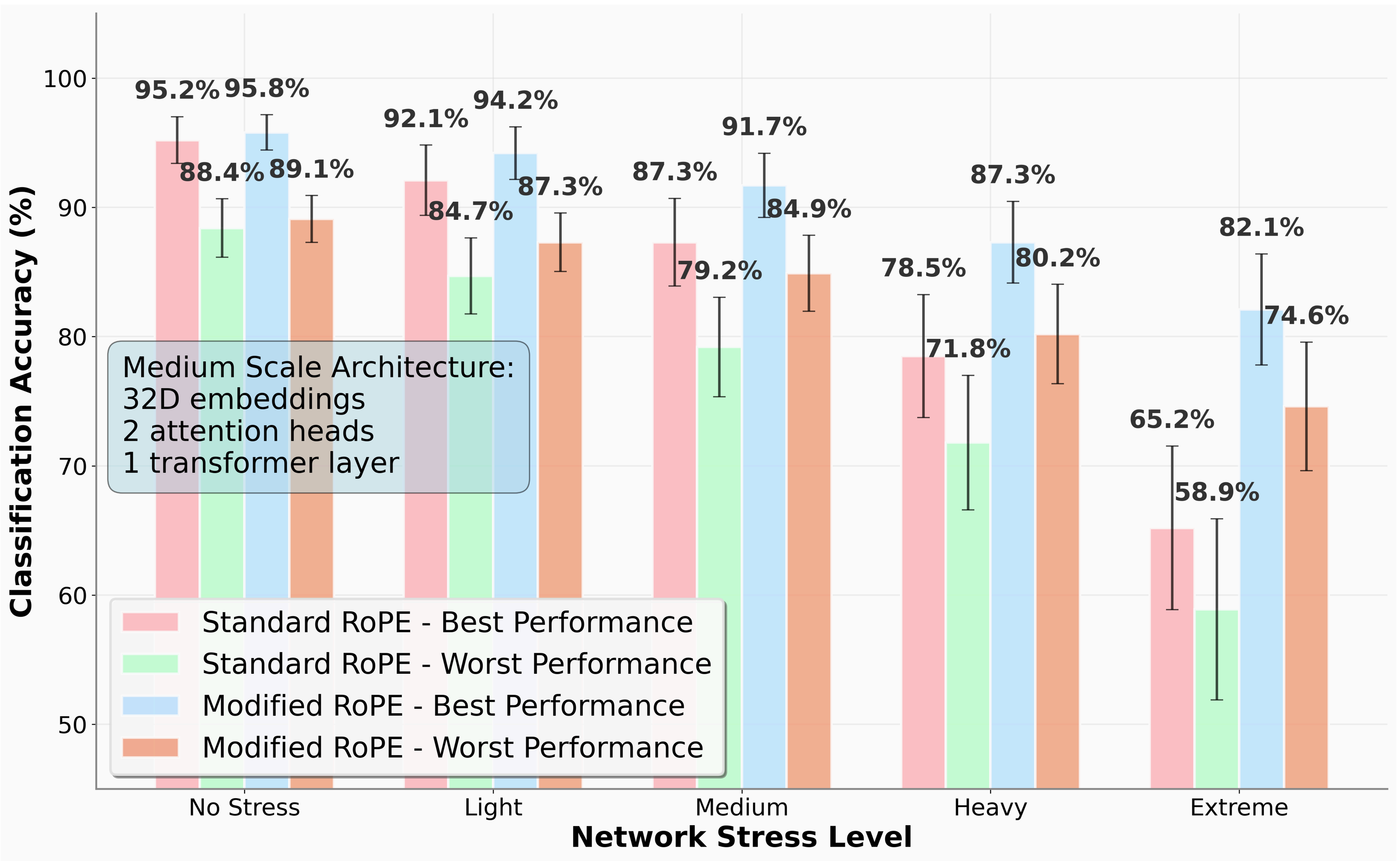
$$\phi_{\text{opt}} = \frac{\sigma^2}{(\mathbb{E}[\gamma(\xi)] - 1)^2 + \sigma^2}$$

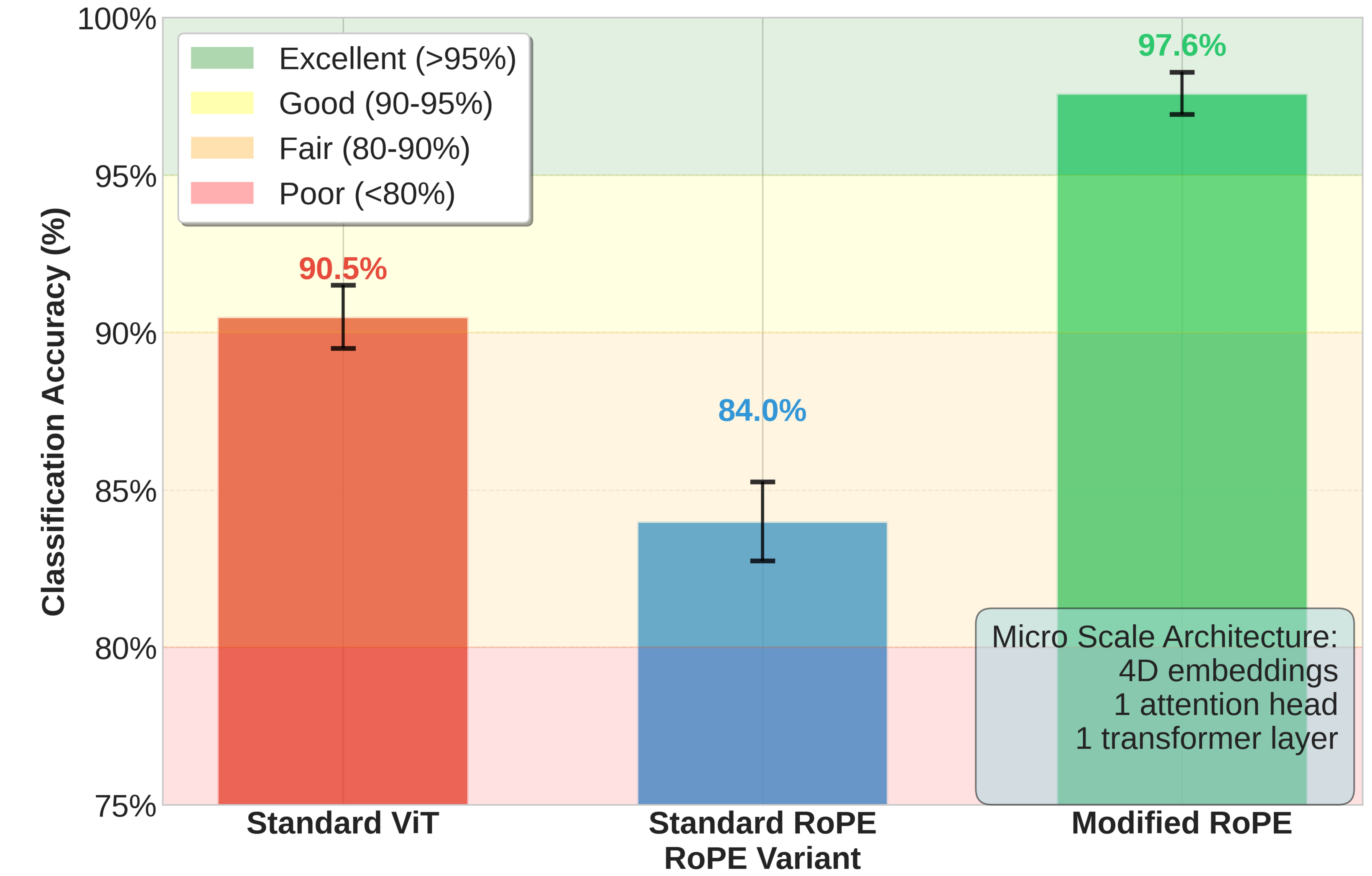
Gap

Existing IoT malware detection methods often struggle under real-world network degradation and device constraints, despite performing well in controlled environments. Although Rotational Positional Encoding can enhance transformer models, its application in IoT deployments is limited due to its inability to adapt to unstable conditions. This work addresses that gap by introducing a regularized, metadata-aware RoPE that dynamically adjusts to network and device status, ensuring robustness while keeping computational costs low

Results

We conducted a comparative evaluation of Modified RoPE, Standard RoPE, and Standard ViT across five network stress levels that simulate real-world scenarios, including WiFi interference and disaster conditions.
The evaluation encompassed conditions from ideal (0% packet loss, 0 ms delay) to extreme degradation (40% packet loss, 200-500 ms delay, 30% bandwidth).





The medium-scale ViT architecture has 32 dimensions for the embeddings and 2 attention heads. Our modified RoPE demonstrates superior robustness, maintaining an accuracy of 82.1% under extreme stress, compared to Standard RoPE's 65.2%.

The micro-scale ViT architecture has 4 dimensions for the embeddings and 1 attention head. The modified RoPE achieved 97.6% accuracy, surpassing Standard ViT's 90.5% and Standard RoPE's 84.0% accuracy even under the worst network condition.

These results demonstrate that enhanced positional encoding significantly improves model performance, even within resource-constrained architectures. Consequently, Modified RoPE is positioned as a critical component for malware detection and other AI applications, as well as edge computing scenarios characterized by network instability and computational limitations.

Personal page

SAINT LOUIS UNIVERSITY

Paper