UNIVERSITAS
GADJAH MADA
**Deteksi dan Mitigasi Silent Data Corruption pada _Deep Neural Network_ Menggunakan Checksum Bobot
dan Penggantian Nilai Aktivasi Berbasis Distribusi Aktivasi Per-Channel**
Davin Jason Evan Raharjo, Muhammad Alfian Amrizal, B.Eng., M.I.S., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

DETEKSI DAN MITIGASI *SILENT DATA CORRUPTION* PADA *DEEP NEURAL NETWORK* MENGGUNAKAN *CHECKSUM* BOBOT DAN PENGGANTIAN NILAI AKTIVASI BERBASIS DISTRIBUSI AKTIVASI *PER-CHANNEL*

Oleh

DAVIN JASON EVAN RAHARJO
22/493367/PA/21184

Silent Data Corruption (SDC) pada *Deep Neural Networks* (DNN) dapat memicu penyimpangan aktivasi yang merambat dan pada akhirnya menurunkan akurasi prediksi. Salah satu pendekatan yang digunakan untuk mendeteksi dan menangani kesalahan tersebut melibatkan pemantauan profil *Distribution of Neuron Activations* (DNA) sebagai referensi aktivasi normal pada kondisi bebas kesalahan. Namun, deteksi berbasis DNA dapat muncul sebelum keluaran melewati bobot yang benar-benar terinjeksi, sedangkan mitigasi berbasis pembatasan aktivasi masih berisiko mengubah representasi fitur yang valid. Oleh karena itu, penelitian ini mengevaluasi deteksi berbasis *checksum* bobot dan strategi mitigasi aktivasi berbasis profil DNA untuk memulihkan prediksi yang terdampak SDC sekaligus menjaga akurasi model.

Melalui rangkaian eksperimen yang diawali dengan pemrofilan keluaran lapisan model pada kondisi normal untuk menetapkan referensi DNA, penelitian ini mensimulasikan kegagalan dengan melakukan *fault injection* pada bobot model. Deteksi penyimpangan selama proses inferensi diidentifikasi menggunakan mekanisme *checksum* untuk memicu tahap mitigasi. Penelitian ini mengevaluasi AF (*Activation Filtering*) sebagai pembatasan keluaran lapisan secara global, D-AF (*DNA-guided Activation Filtering*) sebagai pembatasan berbasis batas lapisan dari profil DNA, AR-Layer (*Activation Redistribution-Layer*) sebagai penggantian keluaran lapisan berbasis histogram lapisan, dan AR-Channel (*Activation Redistribution-Channel*) sebagai penggantian aktivasi berbasis histogram *channel*. Evaluasi dilakukan menggunakan masukan yang identik pada tiga skenario: model bersih (*clean*), model yang terdampak kesalahan tanpa mitigasi, dan model yang terdampak kesalahan dengan mitigasi. Analisis difokuskan pada pemulihan prediksi yang terdampak oleh kesalahan serta stabilitas akurasi model setelah strategi mitigasi diterapkan.

Kata kunci: Silent Data Corruption, Distribution of Neuron Activations, Mitigasi, Deep Neural Network, Fault Injection

UNIVERSITAS
GADJAH MADA
**Deteksi dan Mitigasi Silent Data Corruption pada _Deep Neural Network_ Menggunakan Checksum Bobot
dan Penggantian Nilai Aktivasi Berbasis Distribusi Aktivasi Per-Channel**
Davin Jason Evan Raharjo, Muhammad Alfian Amrizal, B.Eng., M.I.S., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

DETECTION AND MITIGATION OF SILENT DATA CORRUPTION IN DEEP NEURAL NETWORKS USING WEIGHT CHECKSUMS AND PER-CHANNEL ACTIVATION DISTRIBUTION-BASED VALUE REPLACEMENT

By

DAVIN JASON EVAN RAHARJO
22/493367/PA/21184

Silent Data Corruption (SDC) in Deep Neural Networks (DNNs) can induce activation deviations that propagate and eventually degrade prediction accuracy. One approach used to detect and handle such errors involves monitoring the Distribution of Neuron Activations (DNA) profile as a reference for normal activation under fault-free conditions. However, DNA-based detection may occur before the output passes through the weights that were actually corrupted, while activation-bounding mitigation can still alter valid feature representations. Consequently, this study evaluates weight-checksum-based detection and DNA-profile-based activation mitigation strategies for recovering SDC-affected predictions while preserving model accuracy.

Through a series of experiments beginning with model layer outputs profiling under normal conditions to establish a DNA reference, this research simulates faults by performing fault injection on model weights. Detection of deviations during inference is identified using a checksum mechanism to trigger the mitigation phase. This study evaluates AF (Activation Filtering) as global model layer output bounding, D-AF (DNA-guided Activation Filtering) as layer-bound activation bounding from the DNA profile, AR-Layer (Activation Redistribution-Layer) as layer-histogram-based layer output replacement, and AR-Channel (Activation Redistribution-Channel) as channel-histogram-based layer output replacement. The evaluation is conducted on identical inputs across three scenarios: the clean model, the corrupted model without mitigation, and the corrupted model with mitigation. The analysis focuses on the recovery of predictions affected by errors and the stability of model accuracy after the mitigation strategies are applied.

Keywords: Silent Data Corruption, Distribution of Neuron Activations, Mitigation, Deep Neural Network, Fault Injection