Measuring Modality Utilization in Multi-Modal Neural Networks

Saurav Singh[1], Panos P. Markopoulos[2], Eli Saber[3], Jesse D. Lew[4], Jamison Heard[5]

[1,3,5]Rochester Institute of Technology, [2]The University of Texas at San Antonio, [4]New York University

2023 IEEE Conference on Artificial Intelligence

ABSTRACT

Multimodal data can boost machine learning performance

Problem: Multimodal neural nets seem to overly rely on dominant modality; we cannot measure / verify

Proposed solution: New modality utilization metric that quantifies network reliance on individual modalities

Tests: NTIRE-21 (classification), MCubeS (segmentation) datasets

Observed Benefit: Metric indeed increases explainability in multimodal neural networks. Significant potential impact in multimodal data fusion across various application

INTRODUCTION

End-to-end multimodal fusion issues:

- Overemphasis on a single, dominant modality
- Hyperparameter mismatch across modalities

Understanding the extent of network's use of each modality can provide insight into network's operation.

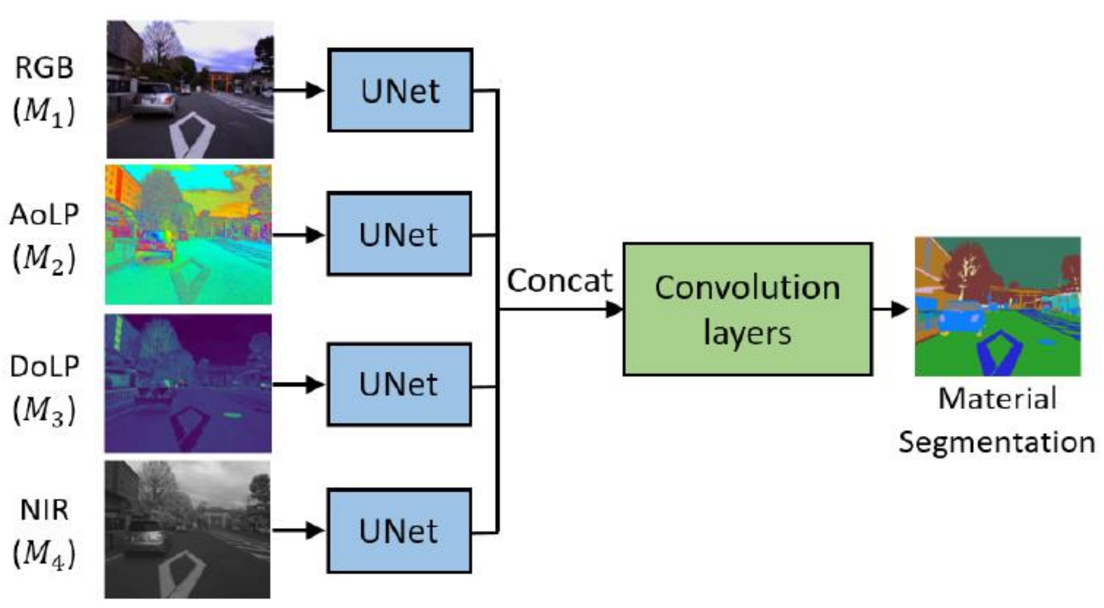

Figure 1: Multimodal Deep Network architecture.

How can we quantify the utilization of a modality by the network?

- Introduced modality utilization (MU) metric.
- Expands permutation feature importance approach to a multi-modal fusion domain.

METHOD

- Method inspired by permutation feature importance [1][2] to compute modality utilization MU_i for each modality
- **Procedure:** Break the association between input modality M_i and output label Y by permuting/shuffling modality M_i samples randomly, leave other modalities $M_j, j \neq i$ unchanged

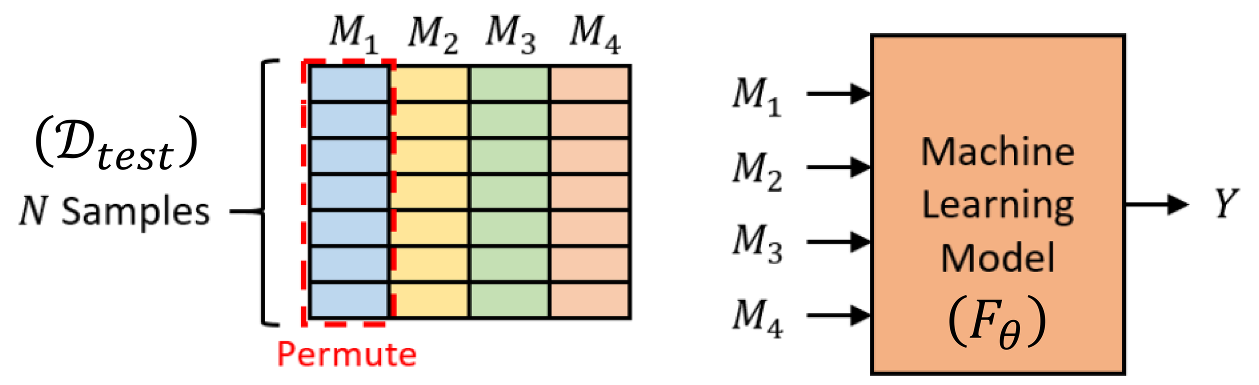

Figure 2: Permuting/shuffling samples of a modality M_i in the dataset to break the association between input modality M_i and the output label Y

Algorithm 1: Compute Modality Utilization

Initialize network model F_θ, and multi-modal test set $\mathcal{D}_{test}$;
Compute model prediction loss L_{test};
for *each modality* M_i **do**
  Randomly permute the samples of modality M_i while keeping the modalities $M_j, j \neq i$ unchanged;
  Compute model prediction loss L_i with permuted modality M_i;
  Compute loss-based Modality Utilization (MU_i) using $MU_i = L_i - L_{test}$;
end
Compute normalized percentages for each MU_i.

RESULTS

Proposed modality utilization metric experimentally assessed on:

- **NTIRE-21** (a classification problem)
- **MCubeS** (an image segmentation problem).

Table I: Performance and modality utilization for the NTIRE-21 dataset

Experiment	Modality	Performance	Modality Utilization (MU) (%)	
		Accuracy (%)	EO	SAR
Unimodal	EO	97.5	100.0	-
Unimodal	SAR	84.9	-	100.0
Multimodal	EO-SAR	97.8	99.59	0.40

Table II: Performance and modality utilization for the MCubeS dataset

Experiment	Modality	Performance	Modality Utilization (MU) (%)			
		mIoU	RBG	AoLP	DoLP	NIR
Unimodal	RGB	0.318	100.0	-	-	-
Unimodal	AoLP	0.266	-	100.0	-	-
Unimodal	DoLP	0.262	-	-	100.0	-
Unimodal	NIR	0.270	-	-	-	100.0
Multimodal	RGB-AoLP-DoLP-NIR	0.374	34.5	19.0	30.9	15.6
Multimodal	AoLP-DoLP-NIR	0.351	-	67.3	21.0	11.7

ABLATION STUDIES ON NTIRE-21 DATASET

- Sensitivity of the modality utilization metric is studied by degrading the quality of the dominant modality (EO)

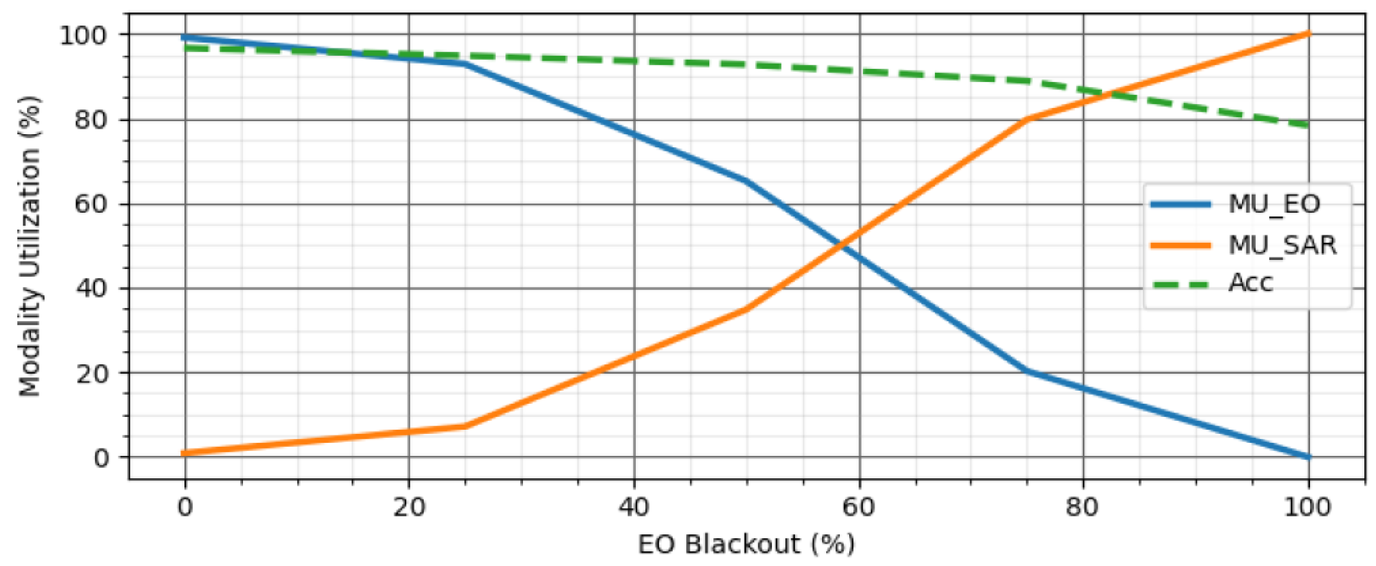

Figure 3: MU and Classification Accuracy by EO Modality Blackout %.

- Evolution of the modality utilization metric while training the network with 50% EO blackout.

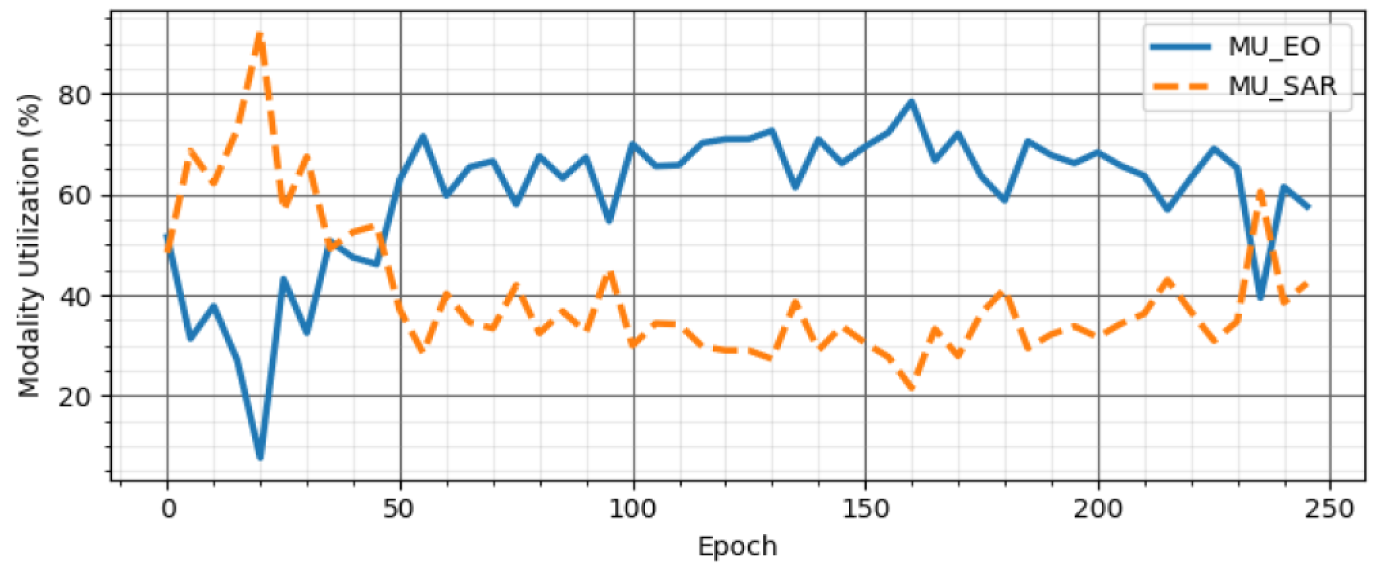

Figure 4: Modality utilization with 50% EO blackout by training epoch.

- Redundancy in modalities impacts network reliance
- Network trained with duplicated EO modalities instead of EO and SAR
- Modality utilization heavily influenced by network's random initialization

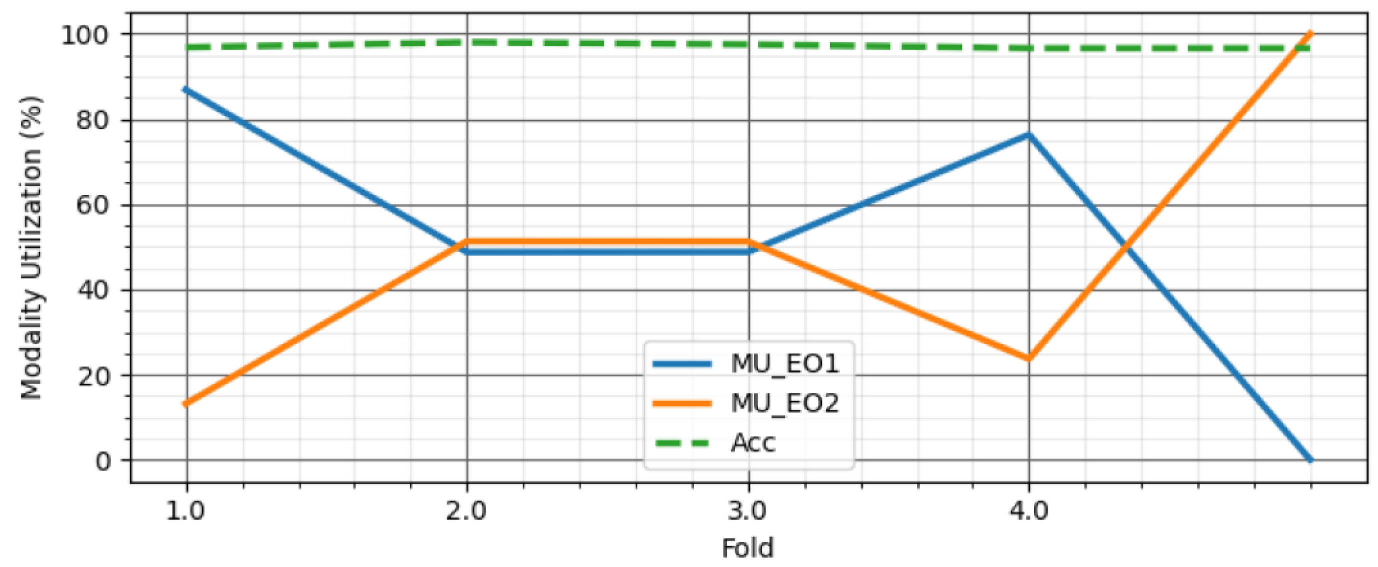

Figure 5: Effects of different network initialization with perfect information redundancy on modality utilization and classification accuracy.

CONTACT INFORMATION

Saurav Singh
ss3337@rit.edu

Panos P. Markopoulos
panagiotis.markopoulos@utsa.edu

Eli Saber
esseee@rit.edu

Jesse D. Lew
jl8429@nyu.edu

Jamison Heard
jrheee@rit.edu

[1] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
[2] A. Fisher, C. Rudin, and F. Dominici, "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously," *Journal of machine learning research: JMLR*, vol. 20, 2019.

This material is based upon work supported in part by the Department of Defense under award HM04761912014 and the Air Force Office of Scientific Research (AFOSR) under award FA9550-20-1-0039.