A Note on 2-simplical attention

by Paul Thompson, 7/4/2025

If you are enjoying reading the new 2-simplicial attention paper [2], it might remind you of Adriel Saporta's amazing paper (below) where she used trilinear forms to find connections across more than 2 data modalities. She extended CLIP to connect audio, video, and text, for cross-modal feature fusion: kind of like cross-attention spanning >2 modalities. If you look at the pink formula, you'll see these are very closely related:

If you're reading the new 2-simplicial paper [2], they try to find useful higher-order intra-sequence correlations, among 3 representations from the same sequence. In Saporta's multi-modal model (which I tweeted about below), she is using trilinear maps across different information sources to fuse them. But you will see something similar in the new paper on 2-simplicial attention, it defines higher-order self-interactions to learn richer token dependencies with a single modality

If you adapted Saporta’s pipeline to fuse 3 modalities with a shared trilinear tensor, you’d get a multi-modal variant of 2-simplicial attention, so you'd be inventing a cross-modal 2-simplicial attention mechanism...

Could this help in medical research? well, with enough data\*, you could learn 3-way interactions, e.g. how combinations of genes + brain signatures predict sets of symptoms, to get at gene->brain->behavior pathways...

You could also write a very tight classifier for brain lesions, e.g. MS lesions (from T1+T2+FLAIR) or cancer (with T1, perfusion MRI + PET), getting at triplets of voxel patterns across sequences (not necessarily co-localized) that together predict pathology, that you could miss w/ pairwise fusion...

You could also use a 2-simplicial or trilinear attention block to fuse 3 modalities simultaneously e.g. to detect gene x environment x brain interactions (hint: are you applying for an autism grant? ;).

Question: Would you need more training data than usual?

Probably, as the parameter count is high (see the pink formula!!). You could add strong priors (as in Li Shen's preference matrix for imaging genetics), to only compute attention on certain triplets? Or LoRA, PEFT [5], etc.

A close-up of a pink background

Description automatically generated

# References

[1] [original Linkedin post by Paul Thompson, 7/4/2025](https://www.linkedin.com/posts/paul-thompson-1380216_if-youre-enjoying-reading-the-new-2-activity-7347034986975285248-vHjM?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAFZfUoBgPoGUucdnvtwuzPv79P8VHj6uvk)

[2] [Fast and Simplex: 2-Simplicial Attention in Triton, Aurko Roy et al, Meta, 2025](https://github.com/dimitarpg13/large_language_models/blob/main/articles/attention/Fast_and_Simplex-2-Simplicial_Attention_in_Triton_Roy_Meta_2025.pdf)

[3] [Note on Multi-Modal Generalization of CLIP by Paul Thompson, 7/3/2025](https://github.com/dimitarpg13/multi-modal_models/blob/main/docs/Note_on_Multi-Modal_Generalization_of_CLIP_by_Paul_Thompson.docx)

[4] [CLIP: connecting text and images, OpenAI blog, Jan 5, 2021](https://openai.com/index/clip/)

[5] [Parameter Efficient Fine-tuning of Transformer-based Masked Autoencoder Enhances Resource Constrained Neuroimage Analysis, NJ Dhinagar et al, 2025](https://github.com/dimitarpg13/multi-modal_models/blob/main/literature/articles/Parameter_Efficient_Fine-tuning_of_Transformer-based_Masked_Autoencoder_Enhances_Resource_Constrained_Neuroimage_Analysis_Dhinagar_2025.pdf)

[6] [Learning Transferable Visual Models From Natural Language Supervision, Alec Radford et al, OpenAI, 2021](https://github.com/dimitarpg13/deep_learning_for_image_processing/blob/main/literature/articles/CLIP/Learning_Transferable_Visual_Models_From_Natural_Language_Supervision_Radford_2021.pdf)

[7] CLIP’s github repo: <https://github.com/openai/CLIP>

[8] [Introducing Triton, OpenAI, July 28, 2021](https://openai.com/index/triton/)

[9] <https://github.com/triton-lang/triton>

[10] <https://triton-lang.org/main/index.html>