* KD Optimal Transport Paper *

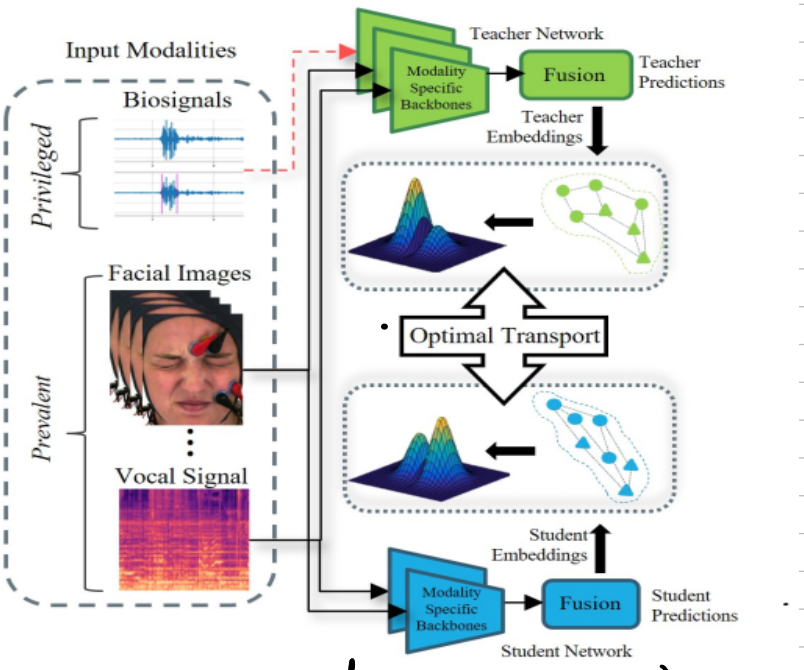
⊙ Multi-modal expression recognition is a problem solved long-ago, in Labs, once these models are put in the wild, they struggle since not all modalities are present at test time (vision, speech, bio-signals), in practice, bio-signals are pretty much absent. (This type of info is called privileged)

⇒ KD have been proposed as a way to distill info from multiple teacher models, each trained on a modality, to one student model. But, KD uses point to point matching ⇒ student learns to match each modality seperately ⇒ has no way of capturing cross-modality structure.

⇒ This paper proposes a way for the student model to capture that structure & aspects that it boosts student performance

⊙ KDOT captures the cross-modality structure by calculating cosine similarity & selecting the top-k anchors (to allow for a flexible distillation process, not rigid)

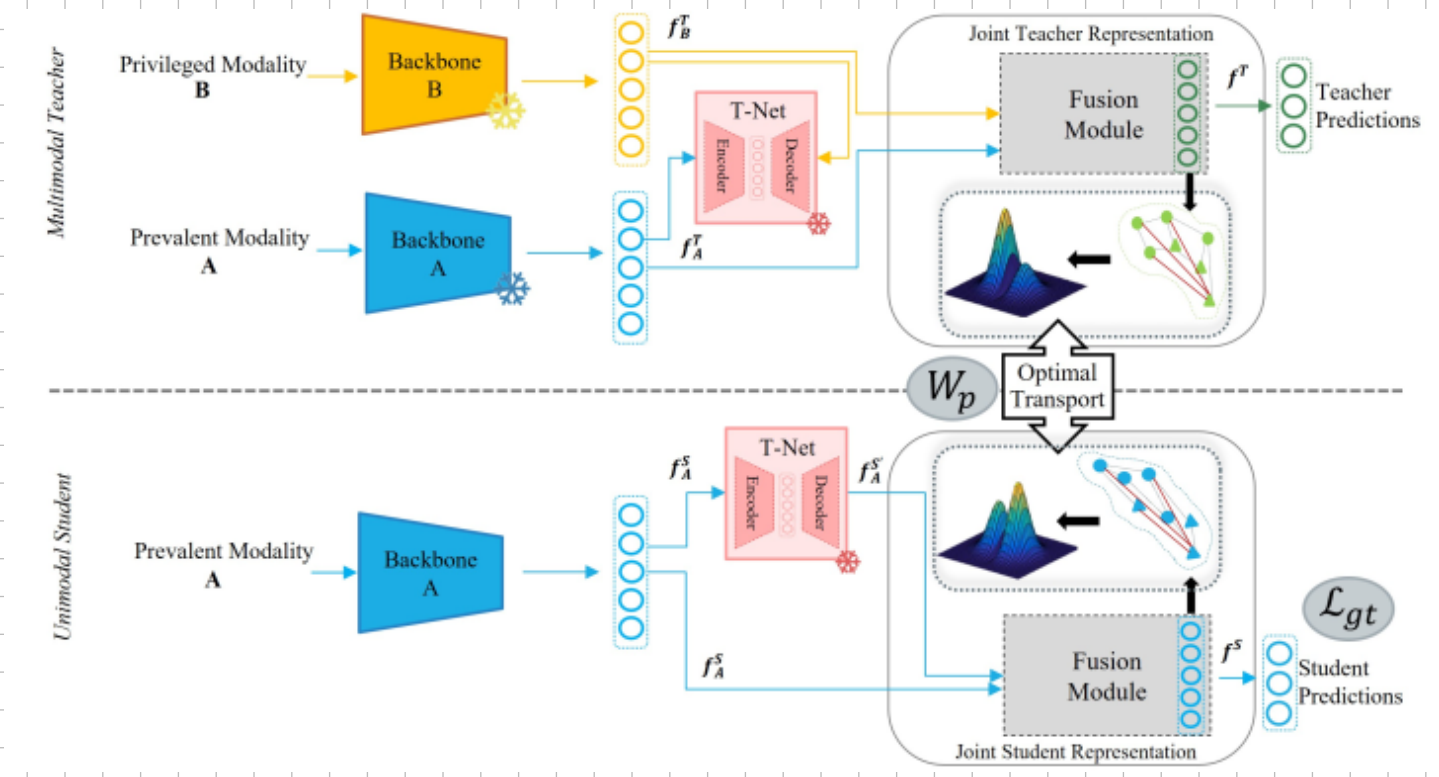
⊙



ditching the previliged modalities (since they exist only in LAB)

① PKDOT's goal is to mitigate the shortcomings privileged modalities introduce to the model (dropped performance at inference time)

② architecture of PKDOT:



. Teacher Network is trained on both prevalent & privileged data to boost performance, & it's trained normally after (Mid-fusion paradigm);

a very interesting block is introduced;
That's the "T-Net", whose role is to map (prevalent, privileged) modalities for each sample, it is trained with the Teacher network & is used "in inference mode" when the distillation is happening, it's role is to hallucinate privileged modalities based on prevalent ones; a good way to create dummy privileged data that's gonna be used to reinforce the cross-modality structure that's gonna be distilled from Teacher to Student.

- Teacher-Student distillation uses OT method & selects k-anchors (most dissimilar samples in the batch) to distill based on. (will be explained thoroughly later).

⊗ structural similarity matching (KNOT + k-anchors)