input_gradient_cosine_similarity
Mean ± Std across pairs