

std(n) for all attention heads for pre-norm

