Cross-modal transfer is helpful to enhance modality-specific discriminative power for scene recognition. To this end, this paper presents a unified framework to i