How Much Can CLIP Benefit Vision-and-Language Tasks?

`Contribution:`

`using CLIP as the visual encoder for diverse V&L tasks.`