## Output task-specific Replaceable Decoder encoded seauence Transformer Encoder Aggregation & **Embedding** per 16x16 patch Input imaae