

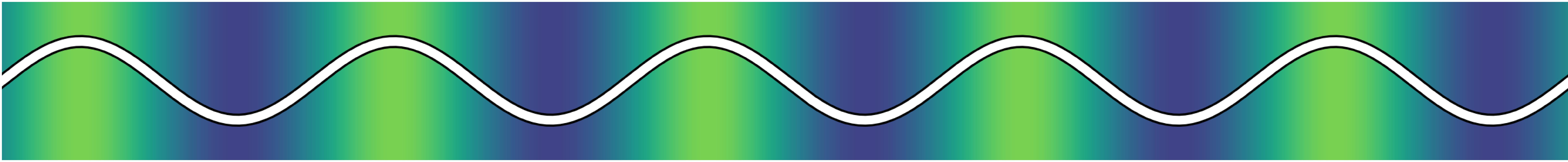


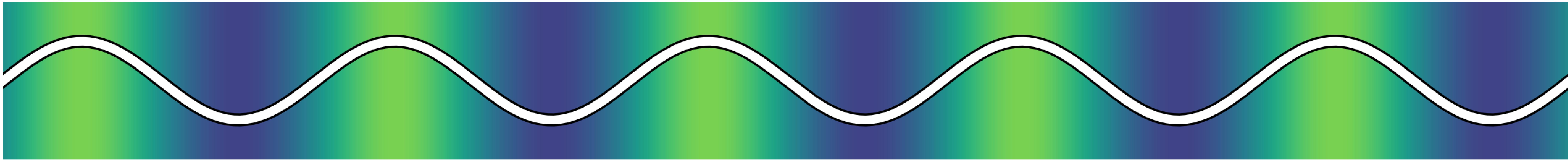




$$v_x(pos, 2i) = \sin \frac{pos}{10000^{\frac{2i}{D}}}$$

$$v_y(pos, 2i + 1) = \cos \frac{pos}{10000^{\frac{2i}{D}}}$$





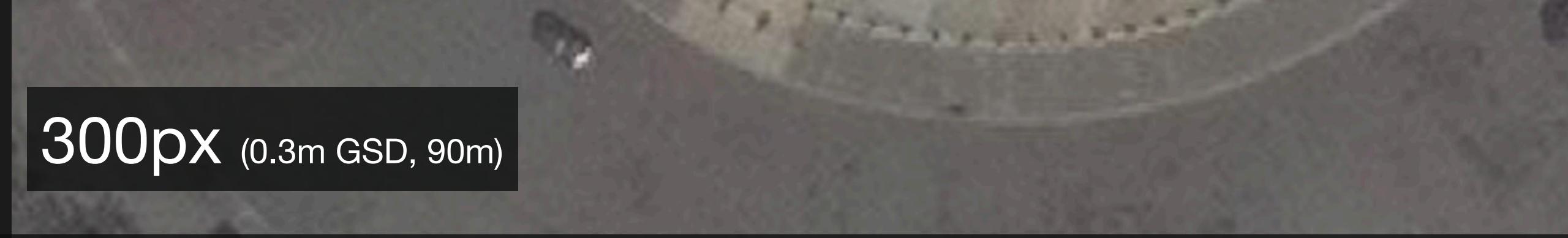
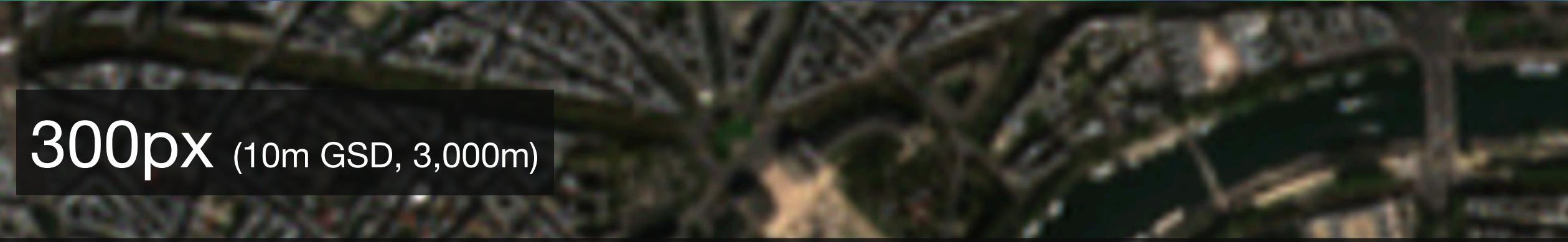
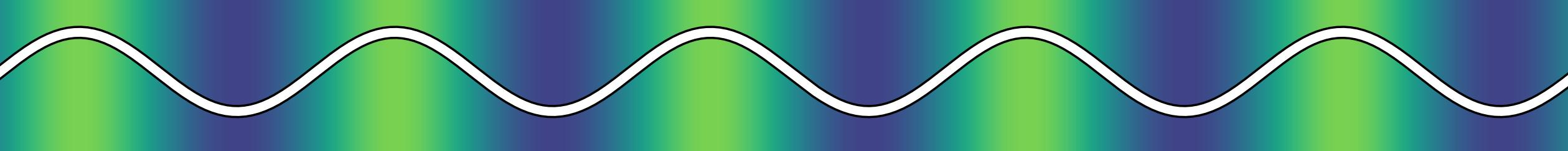
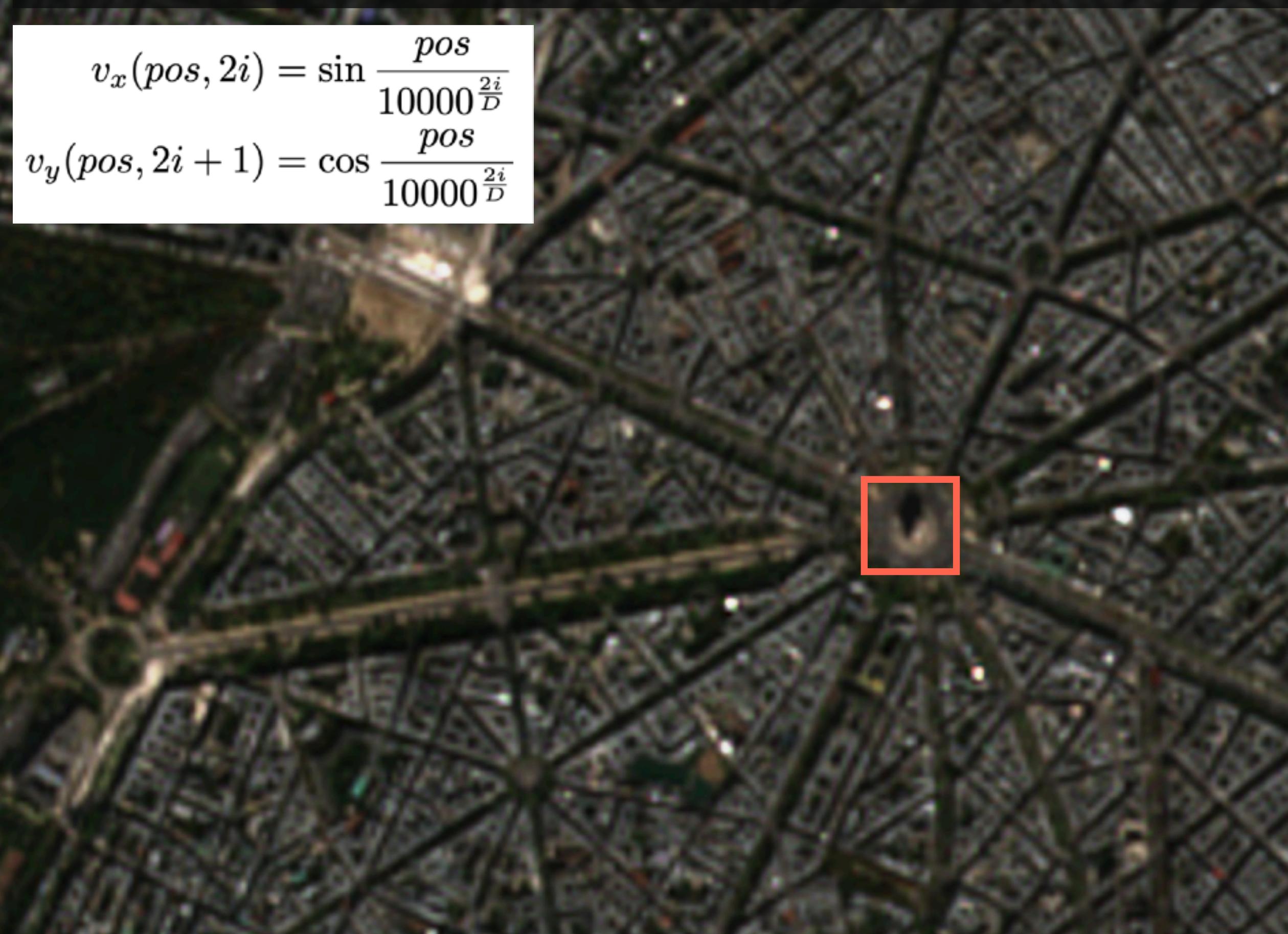
300px (0.3m GSD, 90m)

300px (10m GSD, 3,000m)



Problem: Transformers reason in pixel space

$$v_x(pos, 2i) = \sin \frac{pos}{10000^{\frac{2i}{D}}}$$
$$v_y(pos, 2i + 1) = \cos \frac{pos}{10000^{\frac{2i}{D}}}$$



Problem: Masked autoencoders reconstruct the input

