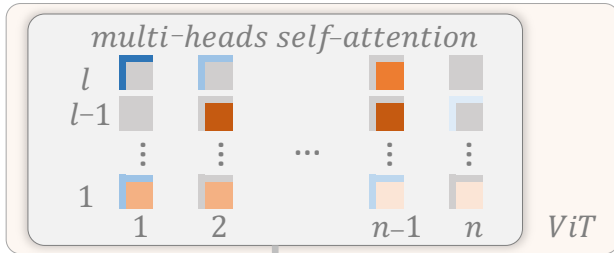


input images



Predictions

all $l \times n$ heads

I. WHAT heads are important?

§5.2

layer & model impact

local & global impact

important heads

II. WHY are they important?

CLS token



patch token



pattern

strength

CLS and patch attention patterns

§5.4



patch attention strength

§5.3

