

Computerized tomographies scan of the head and neck in a **male** patient, aged **29**, weighing **85** kg. The **male** patient **has** undergone chemotherapy.

PatientID	Age	...
CHGJ003	29	...
...	...	...

$1 \times 4$

## Decoder

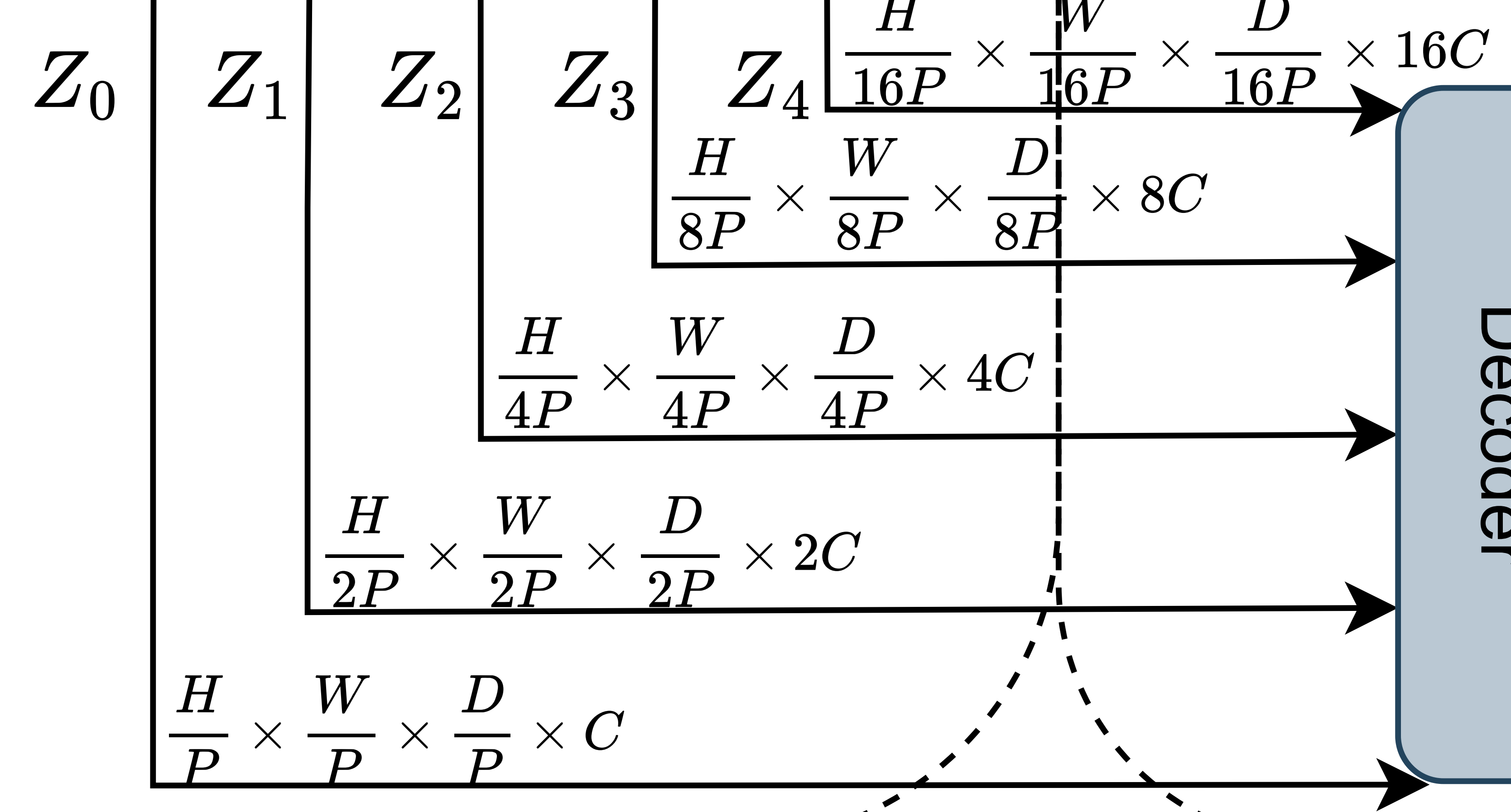


$H \times W \times D \times 2$

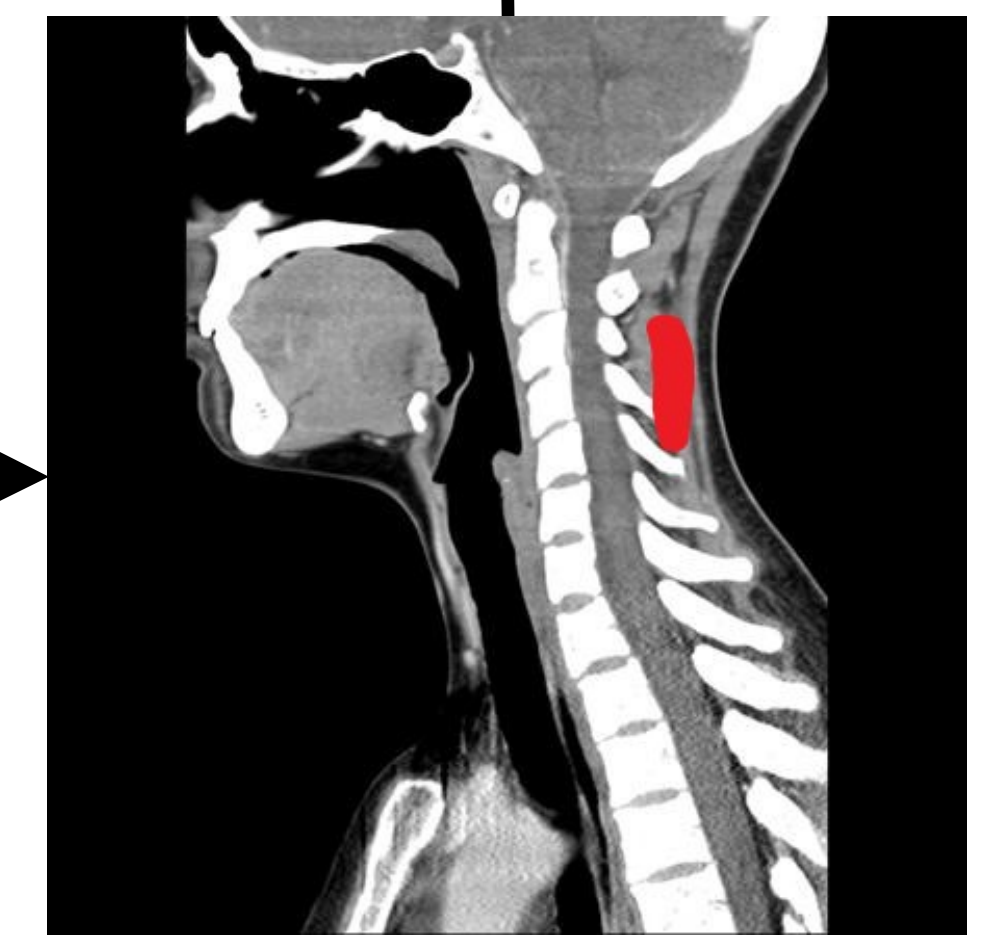
Patch Partition

$\frac{H}{P} \times \frac{W}{P} \times \frac{D}{P} \times C$

Swin Transformer Encoder

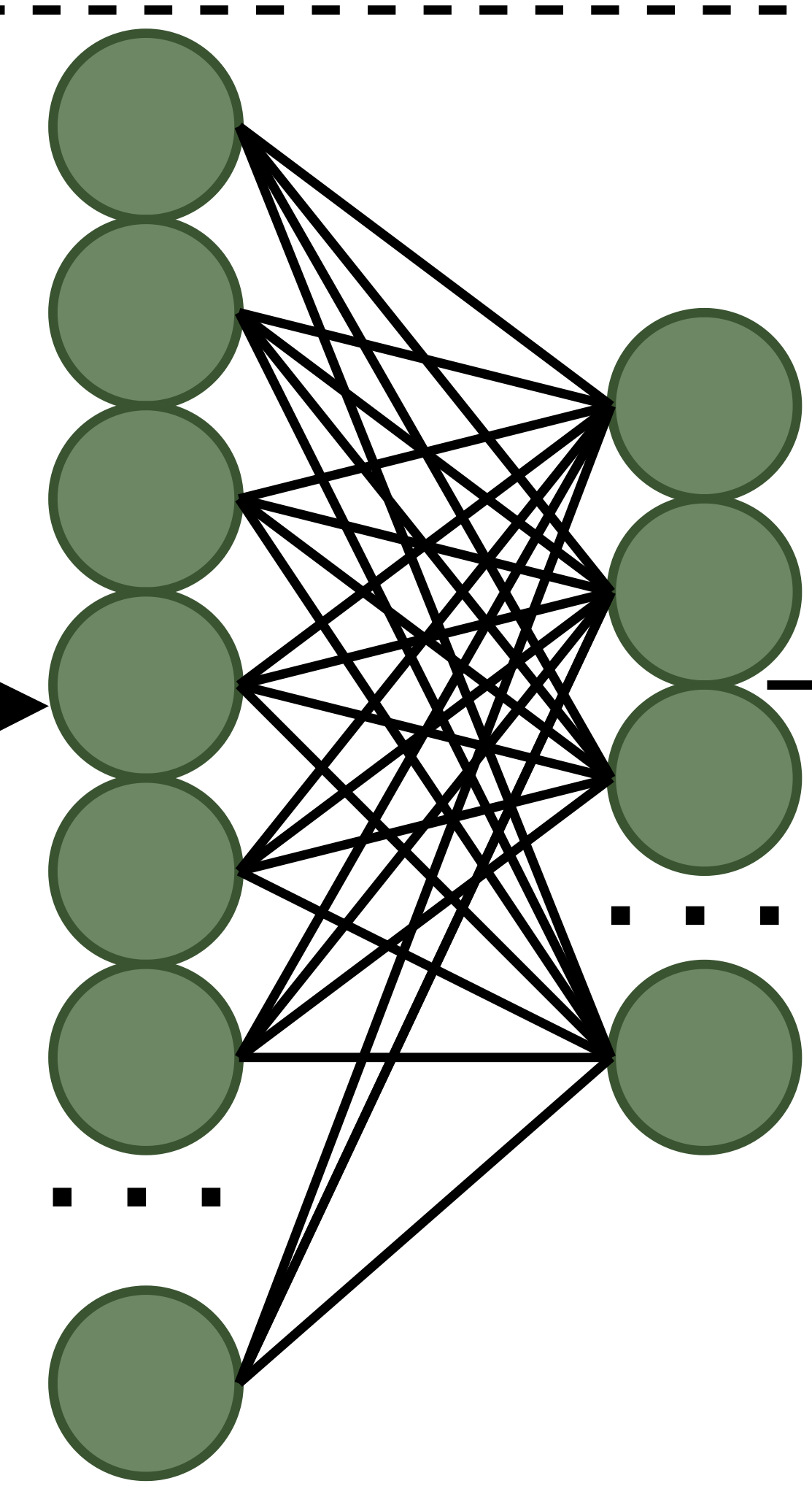


CNN Decoder



Text Encoder

$1 \times d$



$\frac{H}{16P} \times \frac{W}{16P} \times \frac{D}{16P} \times 16C$

Mean

$1 \times 16C$

Mean

$1 \times 16C$

MTLR



$L_{NLL}$

$L_{Final}$

$L_{Dice} + L_{Focal}$