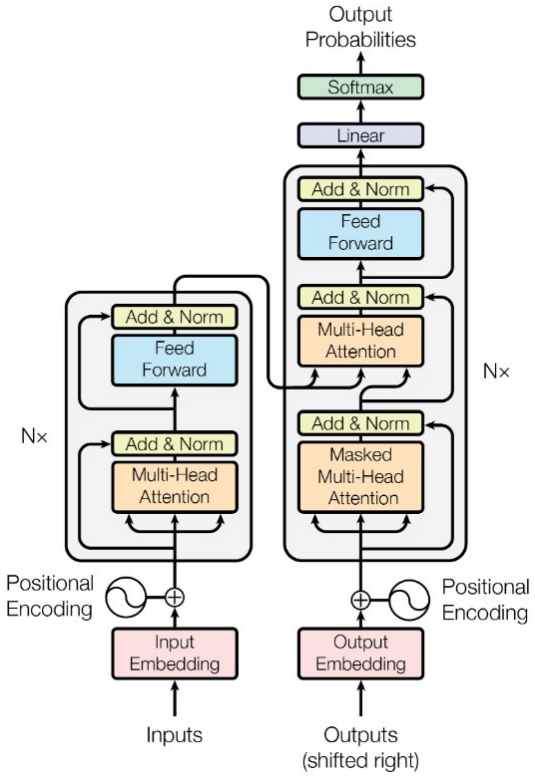
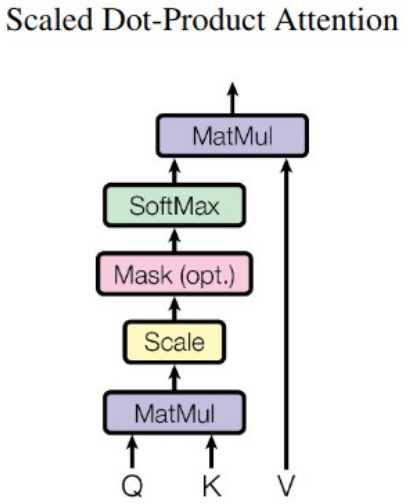
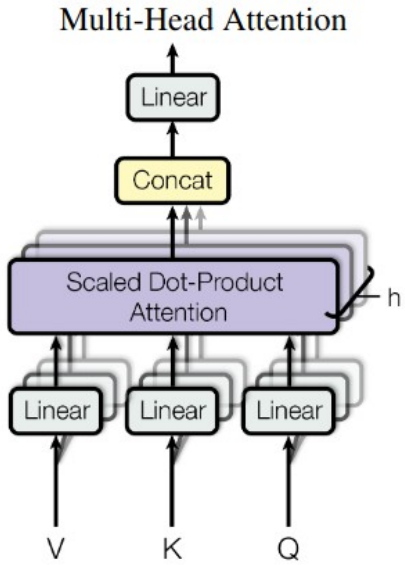
HEXAGON

Visual Transformer

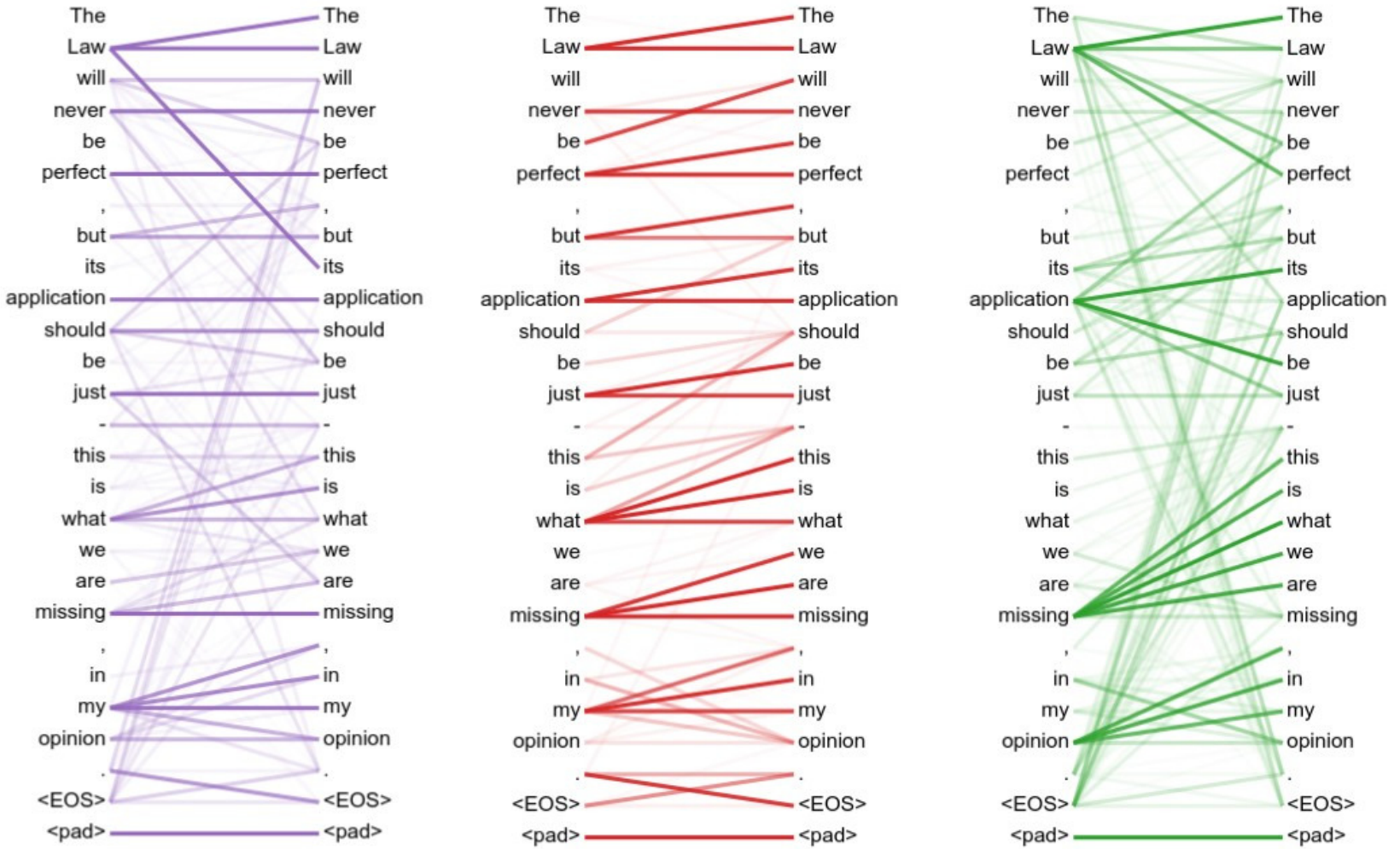
Attention is all you need (Vaswani 2017)





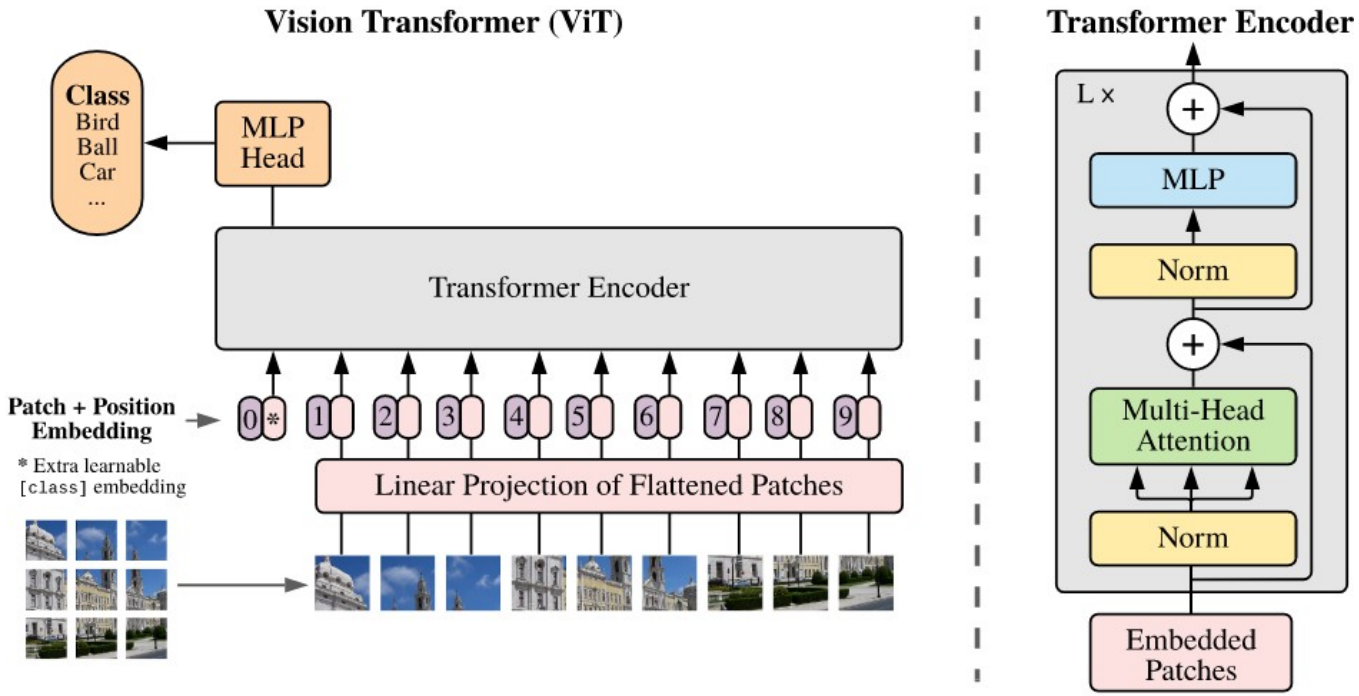


HEXAGON


The
Law
will
never
be
perfect
,
but
its
application
should
be
just
-
this
is
what
we
are
missing
,
in
my
opinion
.
<EOS>
<pad>
The
Law
will
never
be
perfect
,
but
its
application
should
be
just
-
this
is
what
we
are
missing
,
in
my
opinion
.
<EOS>
<pad>

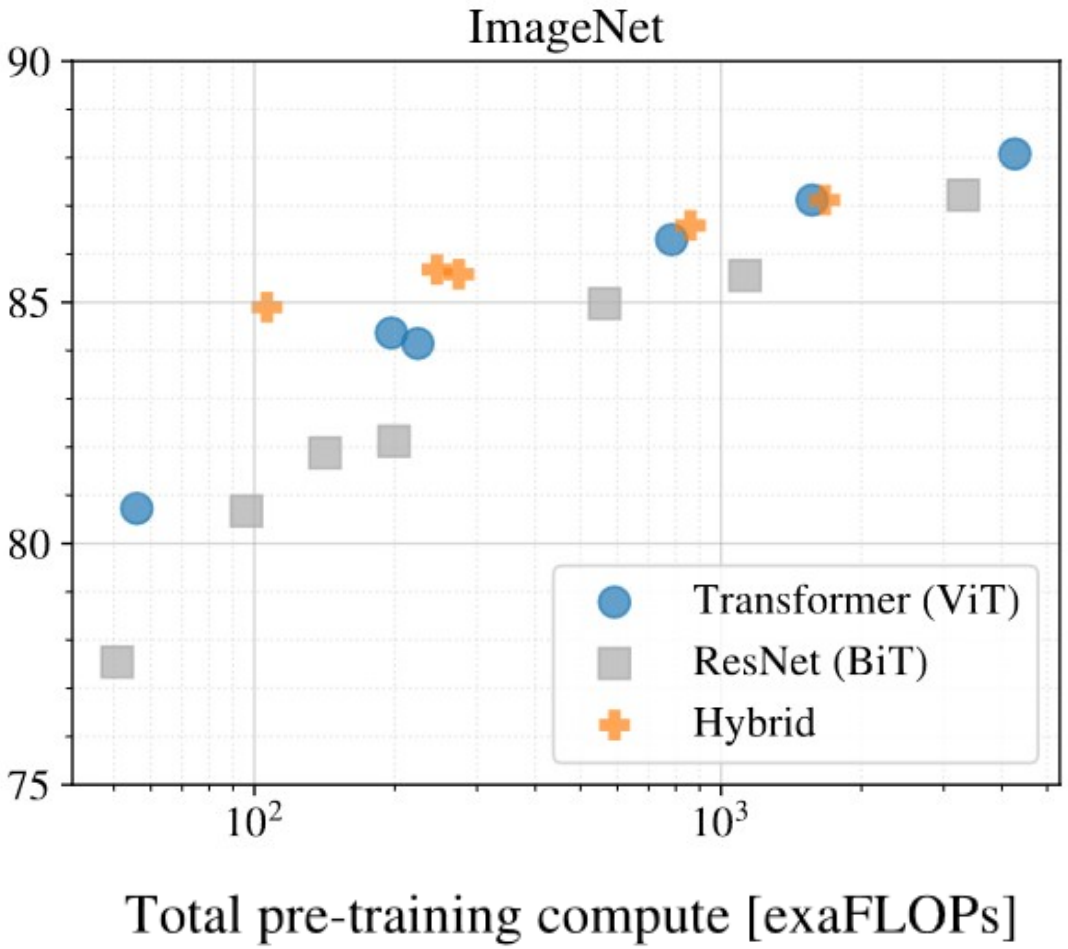
HEXAGON

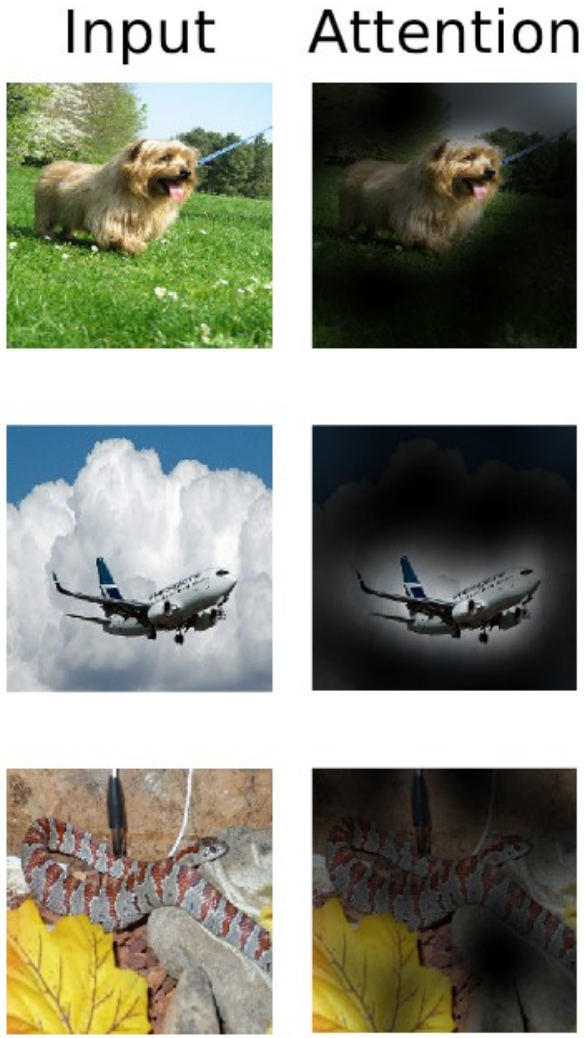
An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (Dosovitskiy 2020)



- Image comme ensemble de patch (16x16)
- Ajout "faux" token pour classif (BERT, Devlin2019)
- 1D conv sur patch
  + position : ordre rasterization

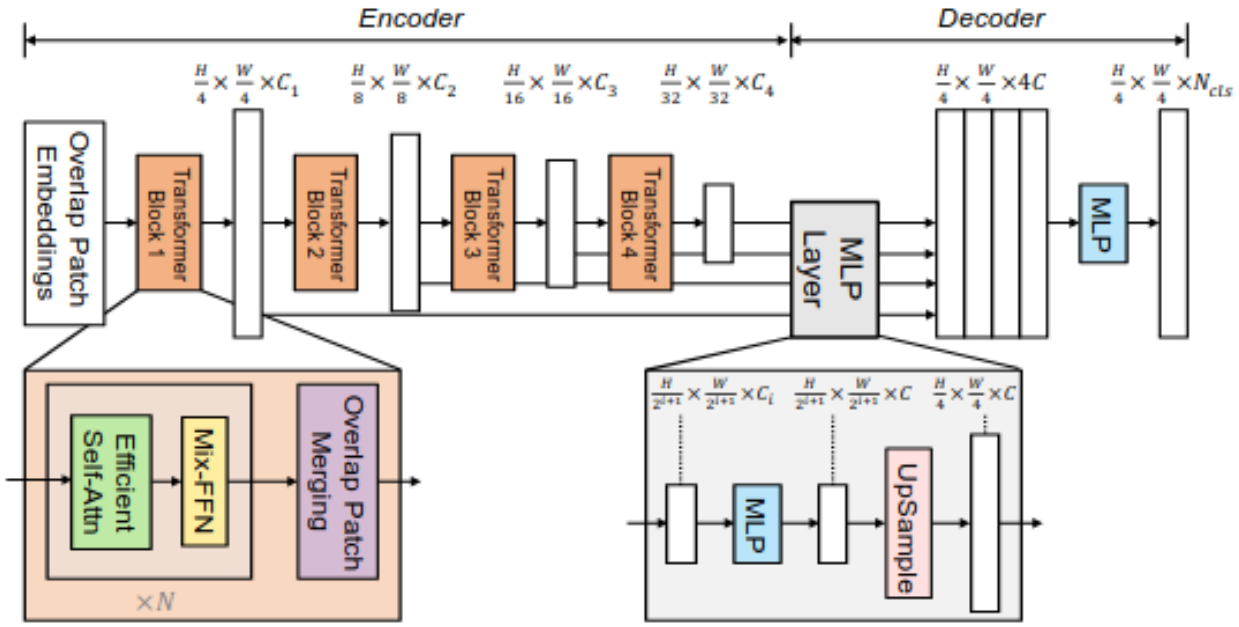
HEXAGON


ImageNet
90
85
80
75
10^2
10^3
Transformer (ViT)
ResNet (BiT)
Hybrid
Total pre-training compute [exaFLOPs]


Input
Attention

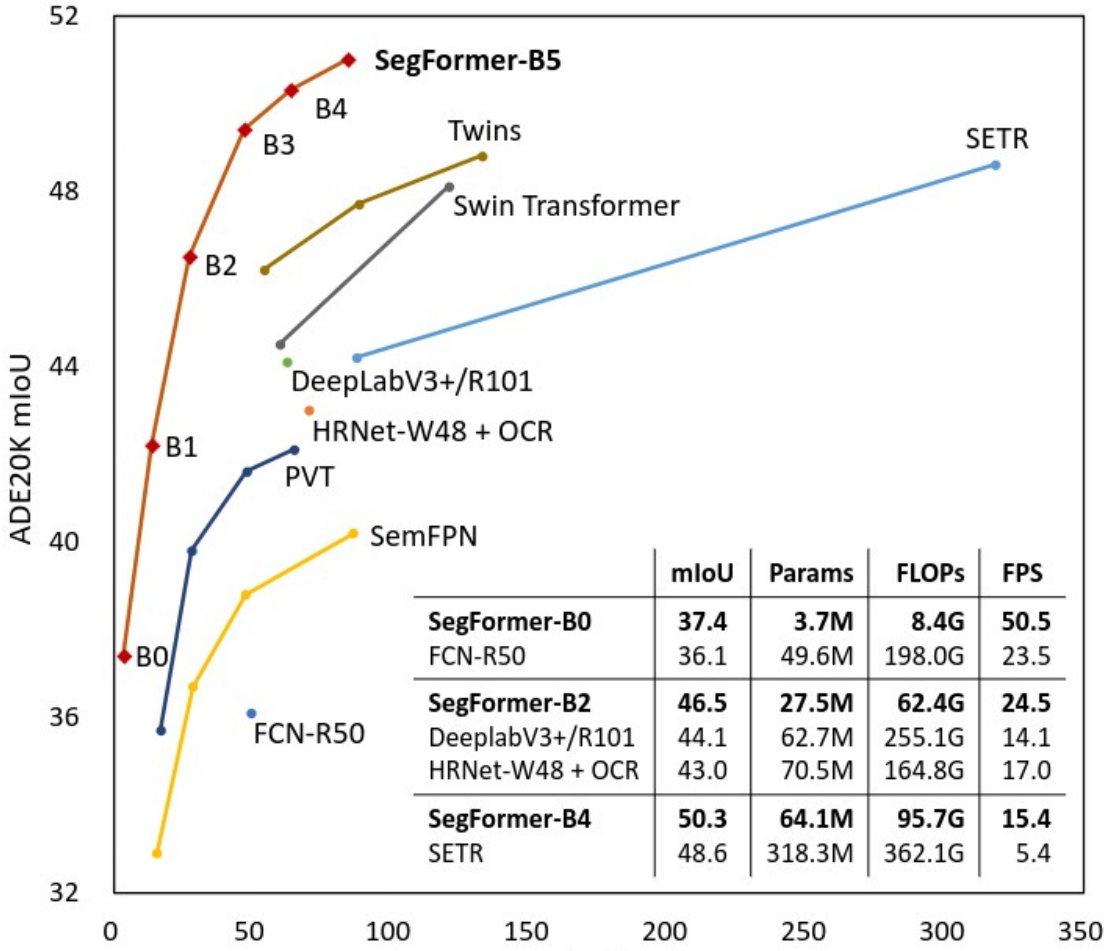
HEXAGON

SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers (Xie 2021)



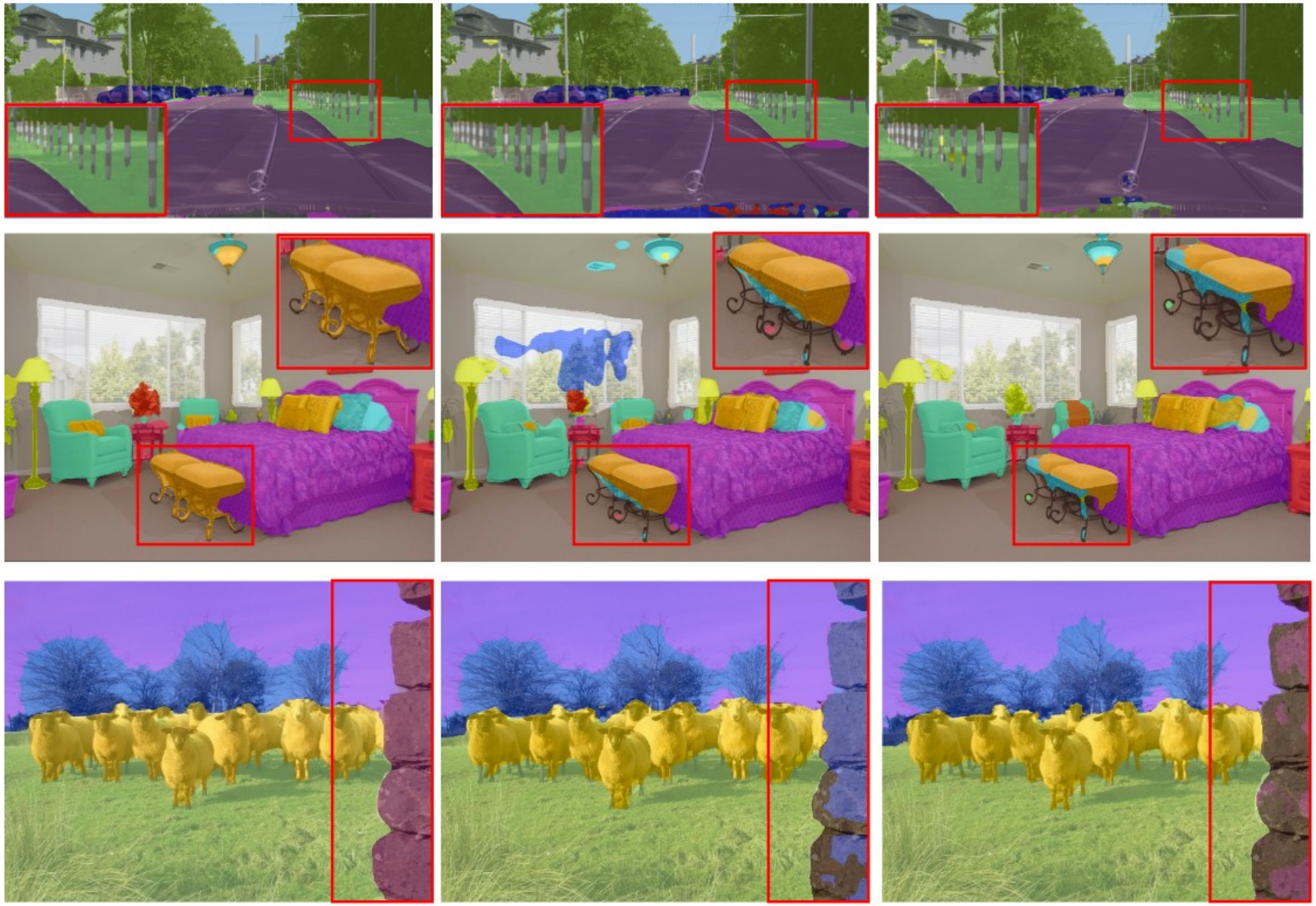
- Retire l'encodage de la position
- Utilise une fenêtre glissante pour extraire les patchs

HEXAGON



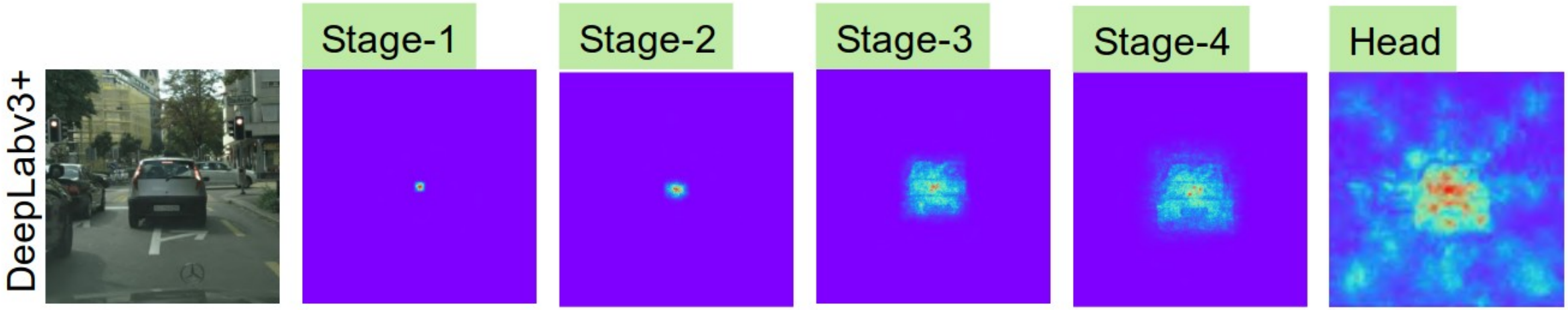
	mIoU	Params	FLOPs	FPS
SegFormer-B0	**37.4**	**3.7M**	**8.4G**	**50.5**
FCN-R50	36.1	49.6M	198.0G	23.5
SegFormer-B2	**46.5**	**27.5M**	**62.4G**	**24.5**
DeeplabV3+/R101	44.1	62.7M	255.1G	14.1
HRNet-W48 + OCR	43.0	70.5M	164.8G	17.0
SegFormer-B4	**50.3**	**64.1M**	**95.7G**	**15.4**
SETR	48.6	318.3M	362.1G	5.4

HEXAGON



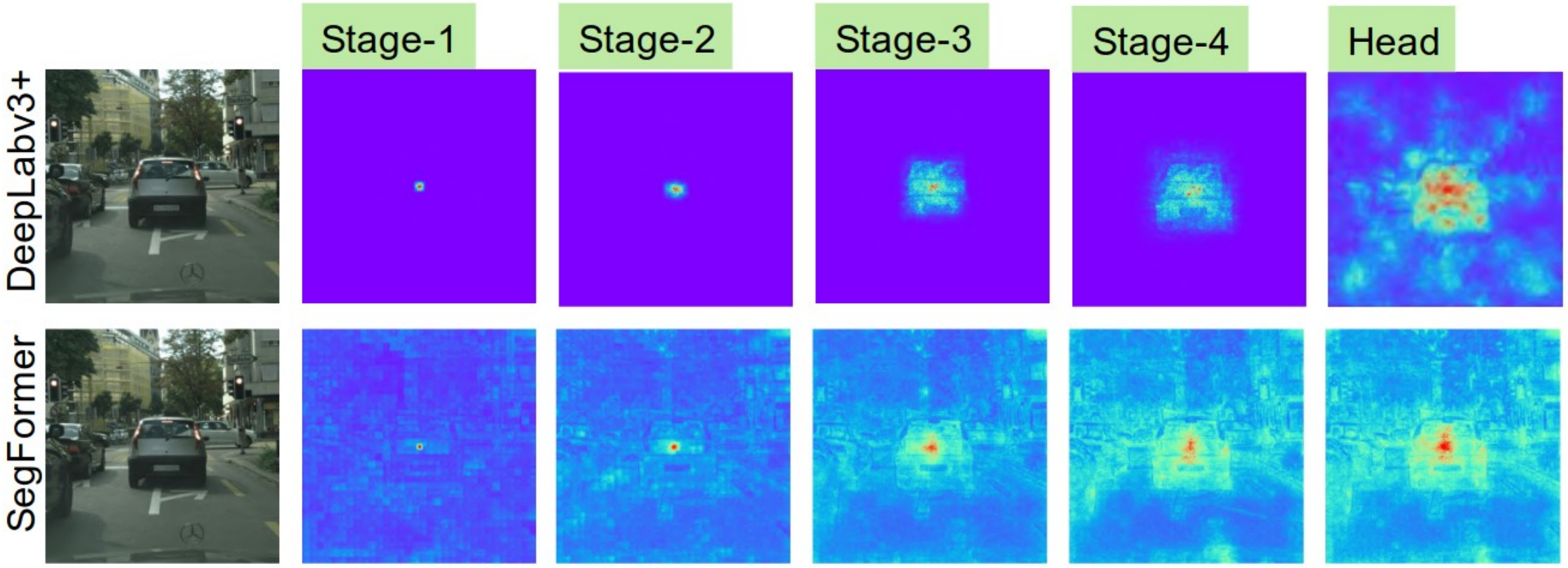
HEXAGON

Pourquoi ?



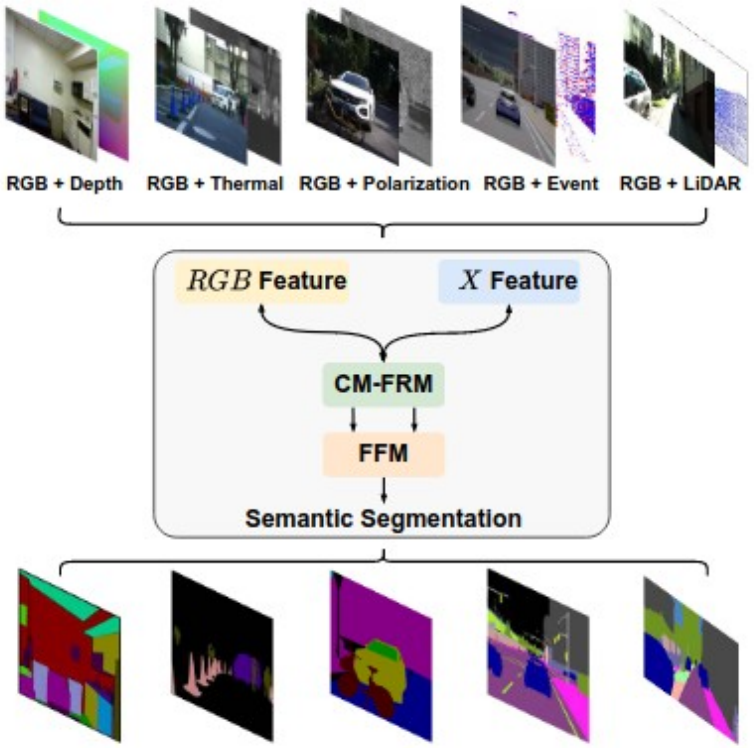
HEXAGON

Pourquoi ?

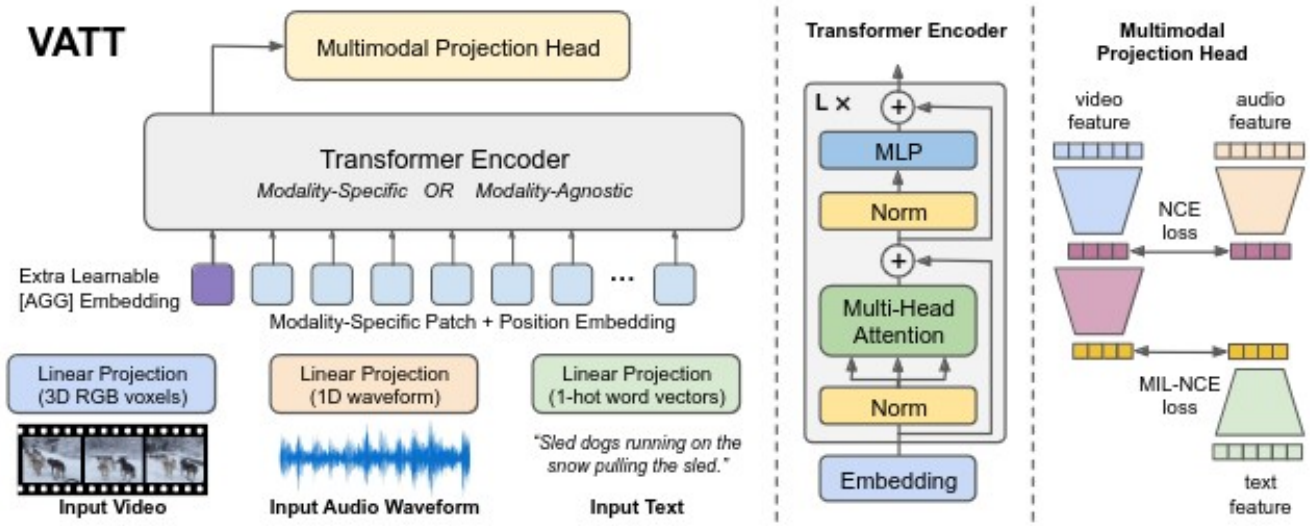


HEXAGON

Approche multimodal



CMX (Zhang 22)



VATT (Akbari 21)

HEXAGON

Ref

"Attention Is All You Need", Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin - NIPS2017

"An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby - ICLR2021

"SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers", Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, Ping Luo - NeurIPS 2021.

"CMX: Cross-Modal Fusion for RGB-X Semantic Segmentation with Transformers", Jiaming Zhang, Huayao Liu, Kailun Yang, Xinxin Hu, Ruiping Liu, Rainer Stiefelhagen

"VATT: Transformers for Multimodal Self-Supervised Learning from Raw Video, Audio and Text", Hassan Akbari, Liangzhe Yuan, Rui Qian, Wei-Hong Chuang, Shih-Fu Chang, Yin Cui, Boqing Gong – NeurIPS 2021