یادگیری عمیق

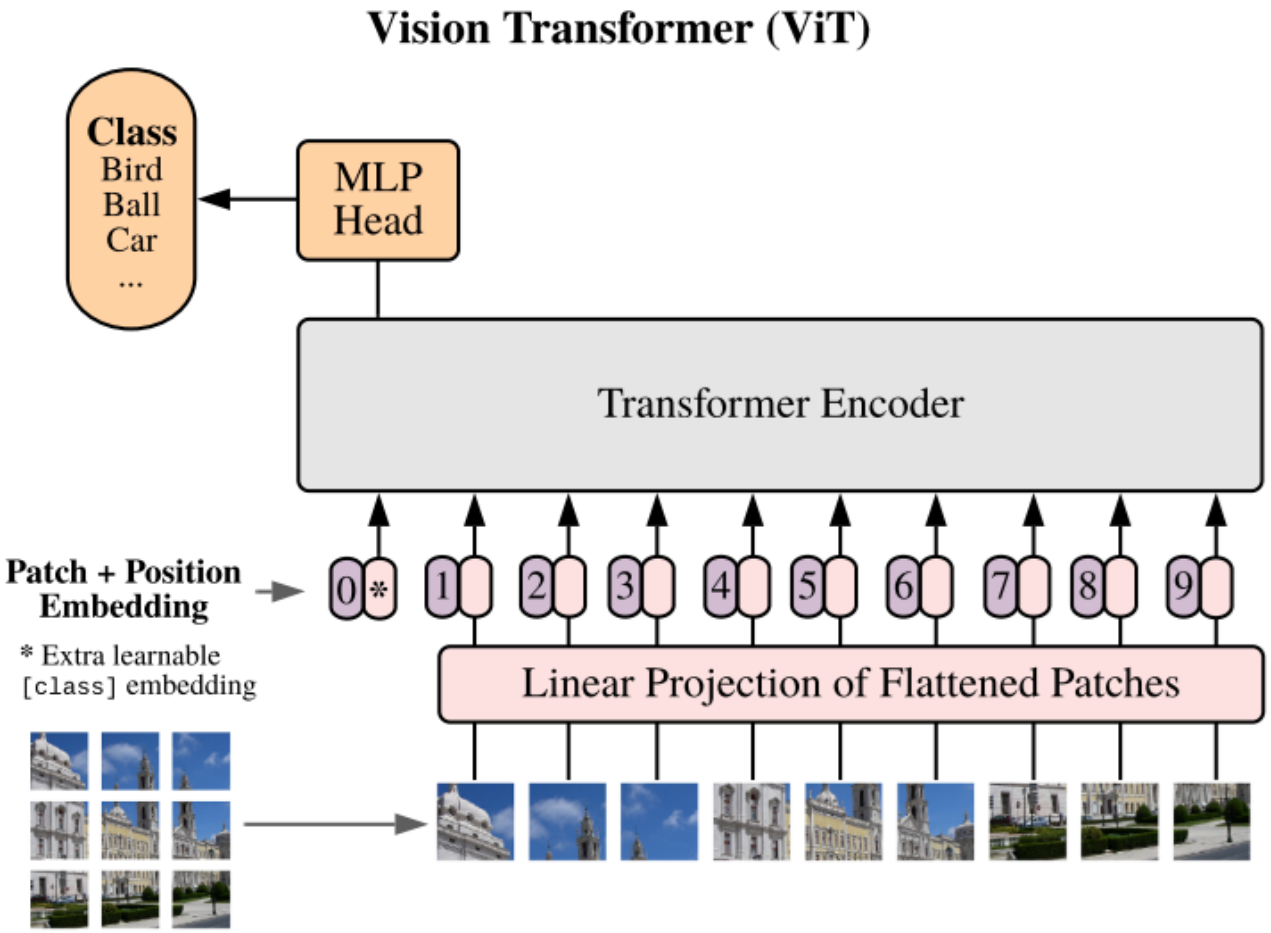
مدرس: محمدرضا محمدی

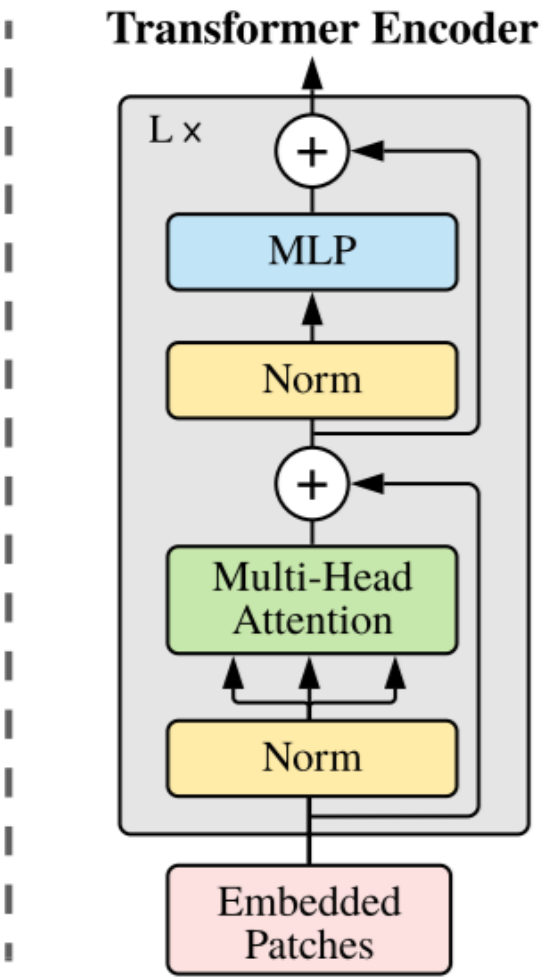
بهار ۱۴۰۲

مکانیزم‌های توجه

Attention Mechanisms

مبدل بینایی (Vision-Transformer)



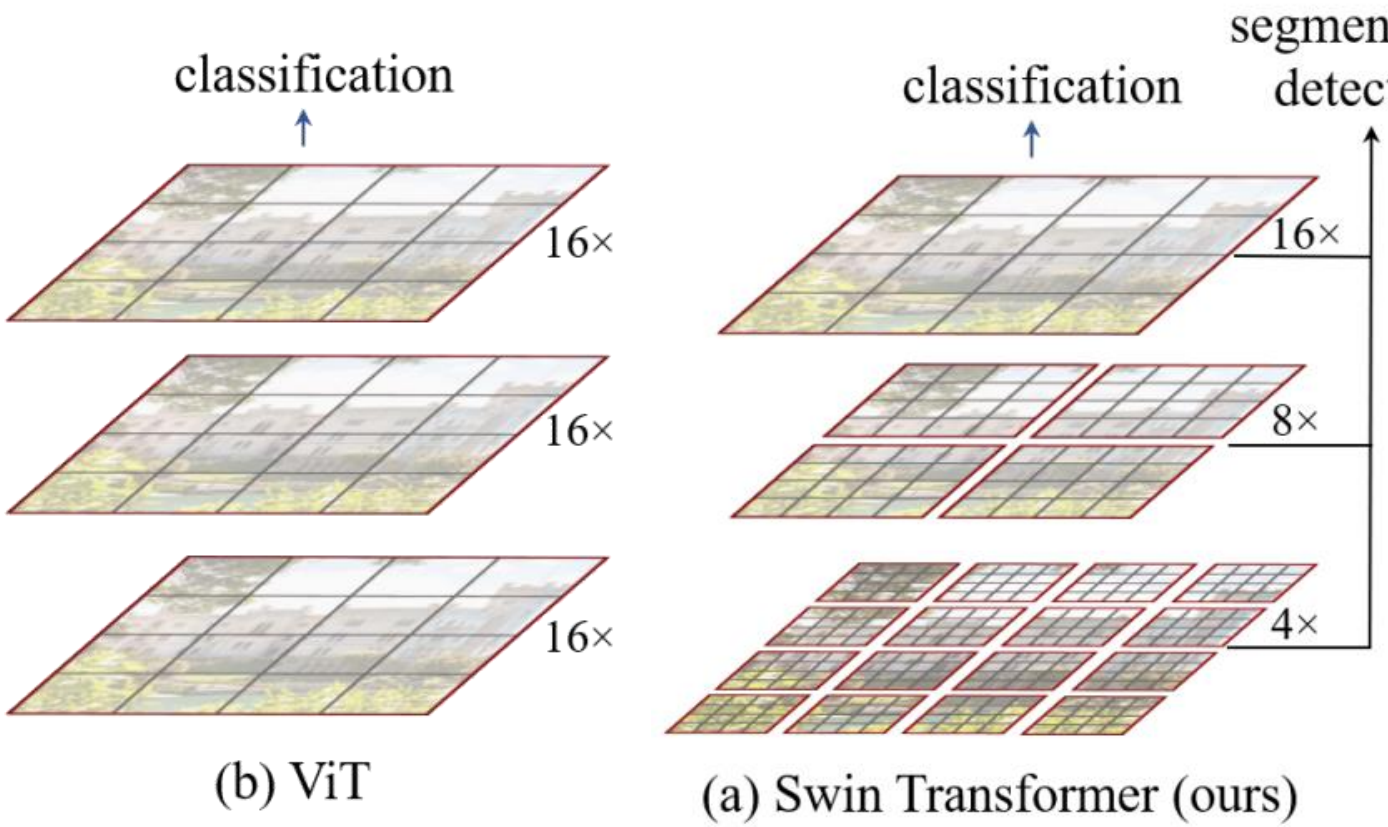


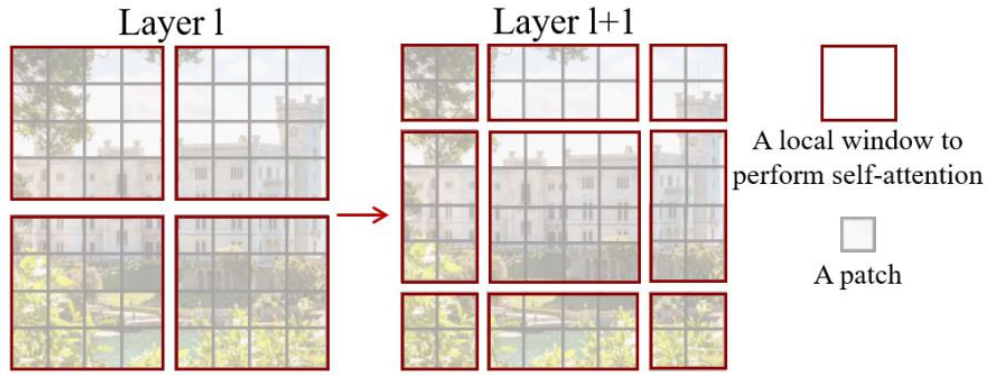
Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv,* 2020..

مبدّل Swin

- در لایه‌های ابتدایی می‌توان تکه‌های کوچکتری از تصویر را پردازش کرد و برای کاهش محاسبات ناحیه توجه به خود را محدود کرد
- در بین لایه‌ها می‌توان محدوده توجه را جابجا کرد

Shifted windows -





Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." CVPR2021.

مبدّل Swin

(a) Regular ImageNet-1K trained models

method	image size	#param.	FLOPs	throughput (image / s)	ImageNet top-1 acc.
RegNetY-4G [44]	224^2	21M	4.0G	1156.7	80.0
RegNetY-8G [44]	224^2	39M	8.0G	591.6	81.7
RegNetY-16G [44]	224^2	84M	16.0G	334.7	82.9
ViT-B/16 [19]	384^2	86M	55.4G	85.9	77.9
ViT-L/16 [19]	384^2	307M	190.7G	27.3	76.5
DeiT-S [57]	224^2	22M	4.6G	940.4	79.8
DeiT-B [57]	224^2	86M	17.5G	292.3	81.8
DeiT-B [57]	384^2	86M	55.4G	85.9	83.1
Swin-T	224^2	29M	4.5G	755.2	81.3
Swin-S	224^2	50M	8.7G	436.9	83.0
Swin-B	224^2	88M	15.4G	278.1	83.5
Swin-B	384^2	88M	47.0G	84.7	84.5

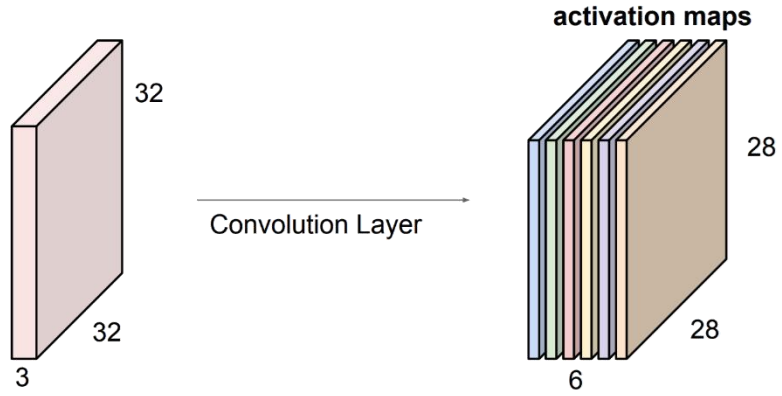
(b) ImageNet-22K pre-trained models

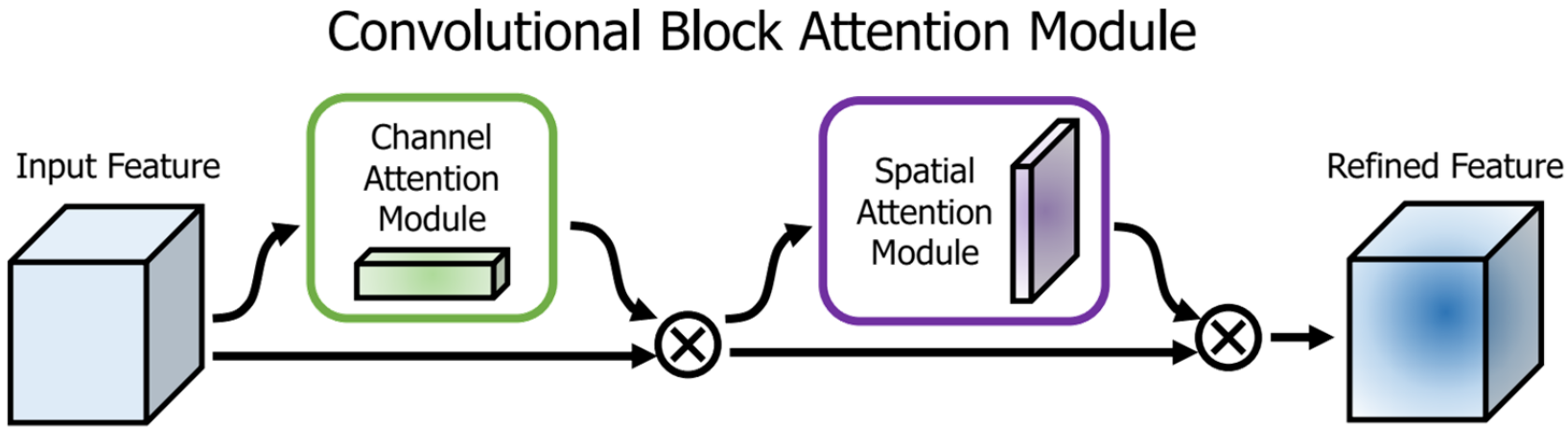
method	image size	#param.	FLOPs	throughput (image / s)	ImageNet top-1 acc.
R-101x3 [34]	384^2	388M	204.6G	-	84.4
R-152x4 [34]	480^2	937M	840.5G	-	85.4
ViT-B/16 [19]	384^2	86M	55.4G	85.9	84.0
ViT-L/16 [19]	384^2	307M	190.7G	27.3	85.2
Swin-B	224^2	88M	15.4G	278.1	85.2
Swin-B	384^2	88M	47.0G	84.7	86.4
Swin-L	384^2	197M	103.9G	42.1	87.3

Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." CVPR2021.

Convolutional Block Attention Module

- برای یک نقشه ویژگی، CBAM به کانال‌ها و مکان‌های بااهمیت توجه می‌کند
  - نقشه‌های توجه که مقادیر آنها در بازه ۰ تا ۱ هستند، در نقشه‌های ویژگی ضرب می‌شوند
  - این دو بخش می‌توانند به ترتیب توجه به چه (what) و کجا (where) را یاد بگیرند
  - بر ویژگی‌های مهم تمرکز می‌شود و موارد غیرضروری تضعیف می‌شوند
  - حذف اطلاعات غیرضروری به بهبود عملکرد مدل کمک می‌کند

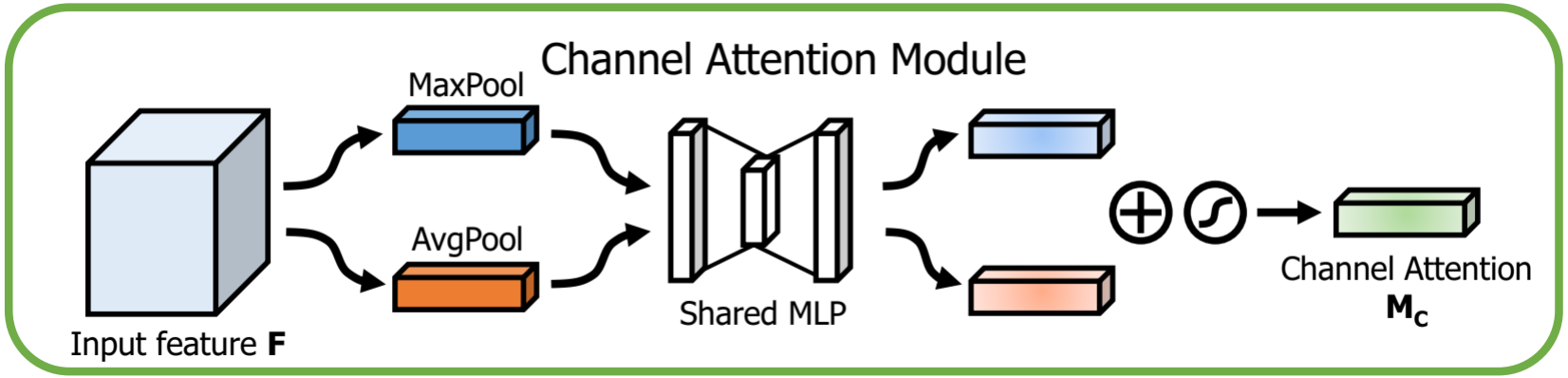


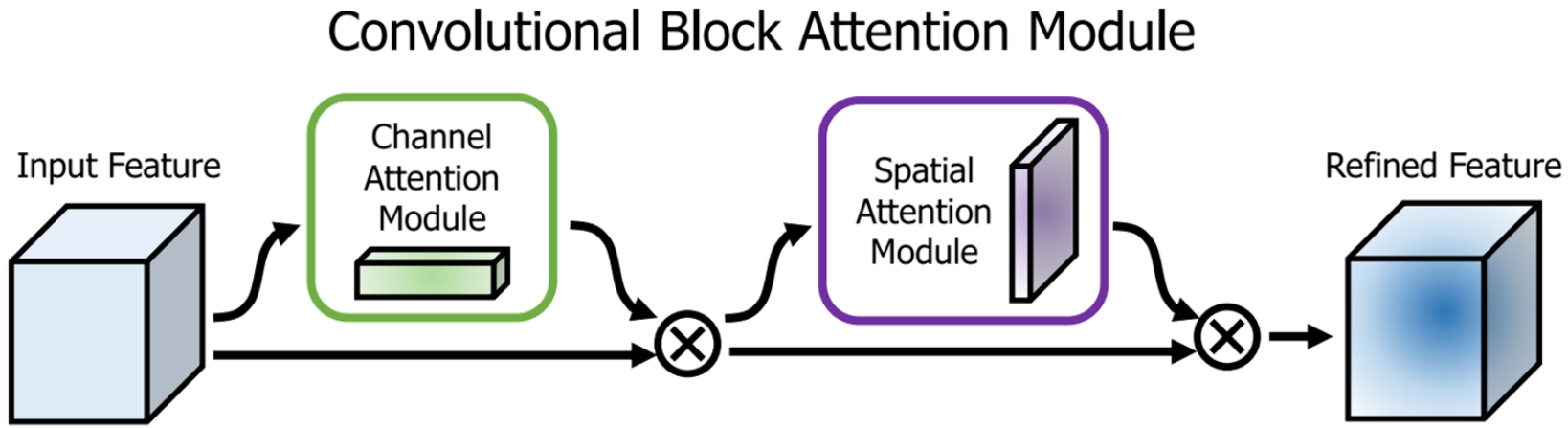


Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." *Proceedings of the European conference on computer vision (ECCV)*, 2018.

توجه کانالی

$$\mathbf{M_c}(\mathbf{F}) = \sigma(\mathbf{W_1}(\mathbf{W_0}(\mathbf{F^c_{avg}})) + \mathbf{W_1}(\mathbf{W_0}(\mathbf{F^c_{max}})))$$



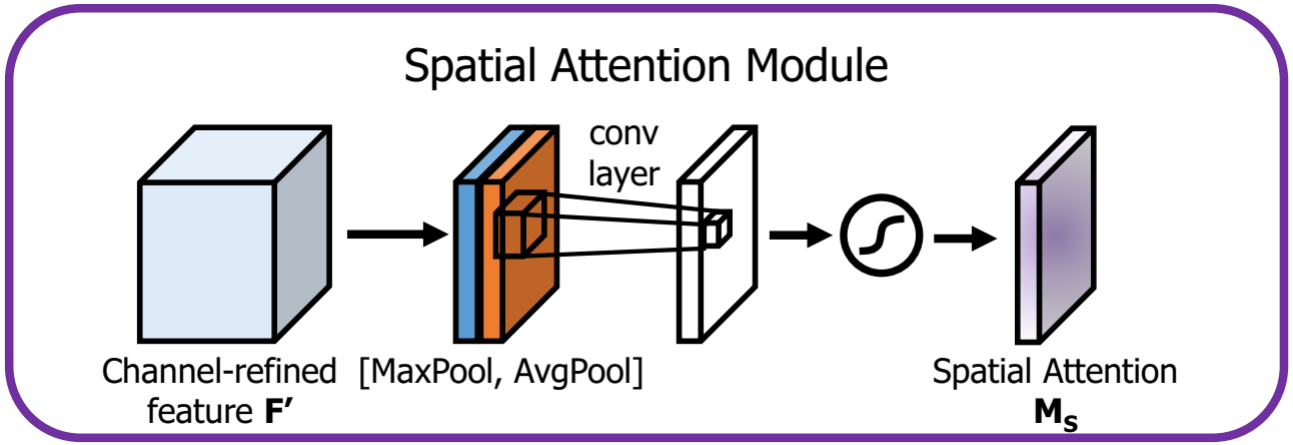


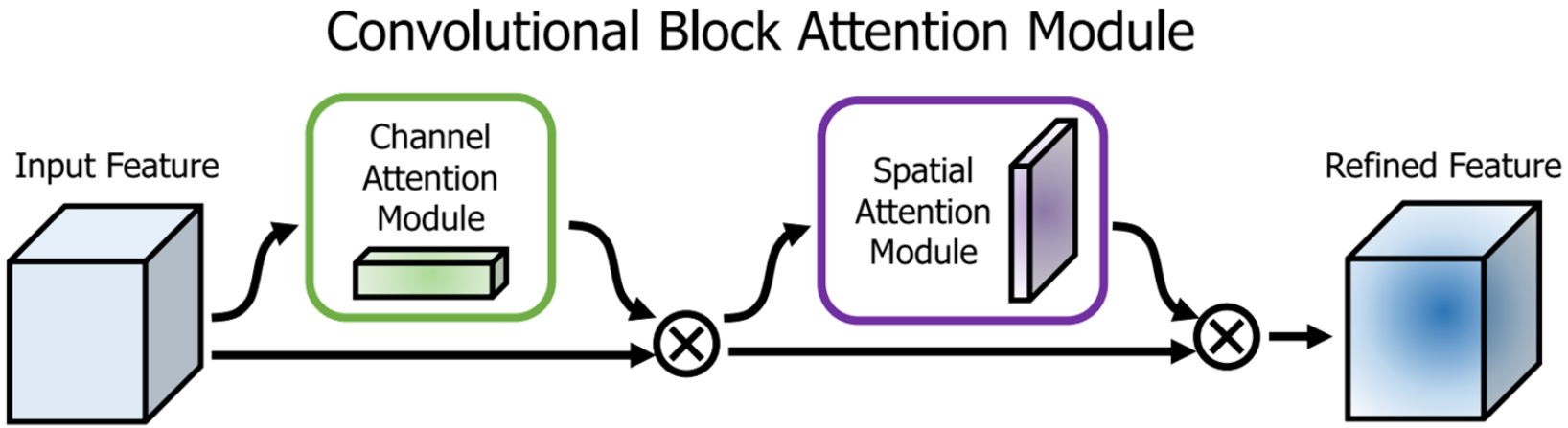
- توجه کانالی بر اینکه چه ویژگی‌هایی با توجه به یک تصویر ورودی معنادار است متمرکز می‌شود
- مقالاتی داریم که مستقیما از خود AvgPool استفاده می‌کنند و هیچ پارامتر قابل آموزشی ندارند اما اینجا از دو نوع ادغام استفاده می‌کند و یک زیرشبکه ۲ لایه را هم آموزش می‌دهد

توجه مکانی

$$\mathbf{M_s}(\mathbf{F}) = \sigma(f^{7\times7}([\mathbf{F}^{\mathbf{s}}_{\mathbf{avg}}; \mathbf{F}^{\mathbf{s}}_{\mathbf{max}}]))$$

- توجه مکانی بر اینکه کدام مکان‌ها با توجه به یک تصویر ورودی معنادار است متمرکز می‌شود





CBAM

Description	Parameters	GFLOPs	Top-1 Error(%)	Top-5 Error(%)
ResNet50 (baseline)	25.56M	3.86	24.56	7.50
ResNet50 + AvgPool (SE [28])	25.92M	3.94	23.14	6.70
ResNet50 + MaxPool	25.92M	3.94	23.20	6.83
ResNet50 + AvgPool & MaxPool	25.92M	4.02	**22.80**	**6.52**

Table 1: **Comparison of different channel attention methods**. We observe that using our proposed method outperforms recently suggested Squeeze and Excitation method [28].

Description	Top-1 Error(%)	Top-5 Error(%)
ResNet50 + channel (SE [28])	23.14	6.70
ResNet50 + channel + spatial	**22.66**	**6.31**
ResNet50 + spatial + channel	22.78	6.42
ResNet50 + channel & spatial in parallel	22.95	6.59

Table 3: **Combining methods of channel and spatial attention**. Using both attention is critical while the best-combining strategy (*i.e.* sequential, channel-first) further improves the accuracy.

CBAM

Architecture	Param.	GFLOPs	Top-1 Error (%)	Top-5 Error (%)
ResNet18 [5]	11.69M	1.814	29.60	10.55
ResNet18 [5] + SE [28]	11.78M	1.814	29.41	10.22
ResNet18 [5] + CBAM	11.78M	1.815	**29.27**	**10.09**
ResNet34 [5]	21.80M	3.664	26.69	8.60
ResNet34 [5] + SE [28]	21.96M	3.664	26.13	8.35
ResNet34 [5] + CBAM	21.96M	3.665	**25.99**	**8.24**
ResNet50 [5]	25.56M	3.858	24.56	7.50
ResNet50 [5] + SE [28]	28.09M	3.860	23.14	6.70
ResNet50 [5] + CBAM	28.09M	3.864	**22.66**	**6.31**
ResNet101 [5]	44.55M	7.570	23.38	6.88
ResNet101 [5] + SE [28]	49.33M	7.575	22.35	6.19
ResNet101 [5] + CBAM	49.33M	7.581	**21.51**	**5.69**
WideResNet18 [6] (widen=1.5)	25.88M	3.866	26.85	8.88
WideResNet18 [6] (widen=1.5) + SE [28]	26.07M	3.867	26.21	8.47
WideResNet18 [6] (widen=1.5) + CBAM	26.08M	3.868	**26.10**	**8.43**
WideResNet18 [6] (widen=2.0)	45.62M	6.696	25.63	8.20
WideResNet18 [6] (widen=2.0) + SE [28]	45.97M	6.696	24.93	7.65
WideResNet18 [6] (widen=2.0) + CBAM	45.97M	6.697	**24.84**	**7.63**
ResNeXt50 [7] (32x4d)	25.03M	3.768	22.85	6.48
ResNeXt50 [7] (32x4d) + SE [28]	27.56M	3.771	**21.91**	6.04
ResNeXt50 [7] (32x4d) + CBAM	27.56M	3.774	21.92	**5.91**
ResNeXt101 [7] (32x4d)	44.18M	7.508	21.54	5.75
ResNeXt101 [7] (32x4d) + SE [28]	48.96M	7.512	21.17	5.66
ResNeXt101 [7] (32x4d) + CBAM	48.96M	7.519	**21.07**	**5.59**

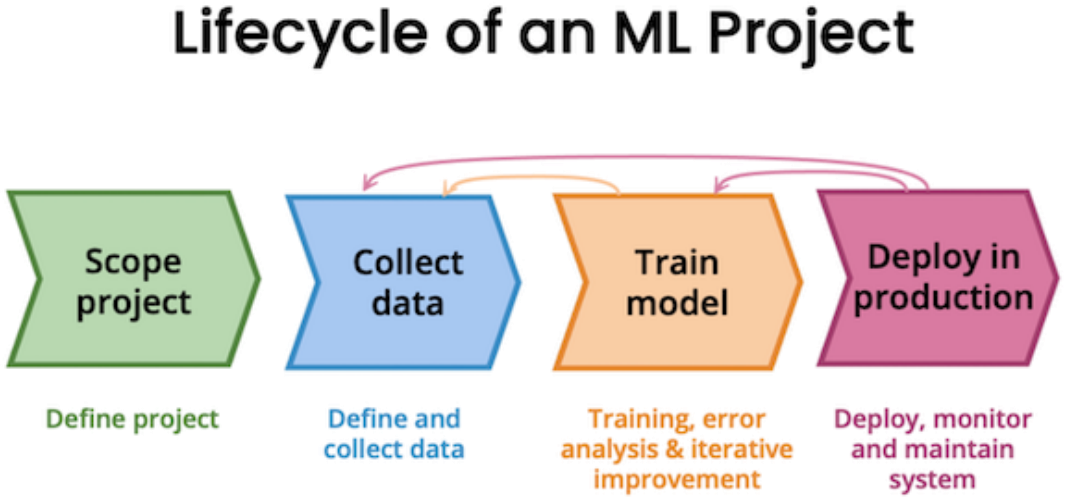
روش‌شناسی کاربردی

Practical Methodology

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning (Chapter 11)*. MIT press, 2016.

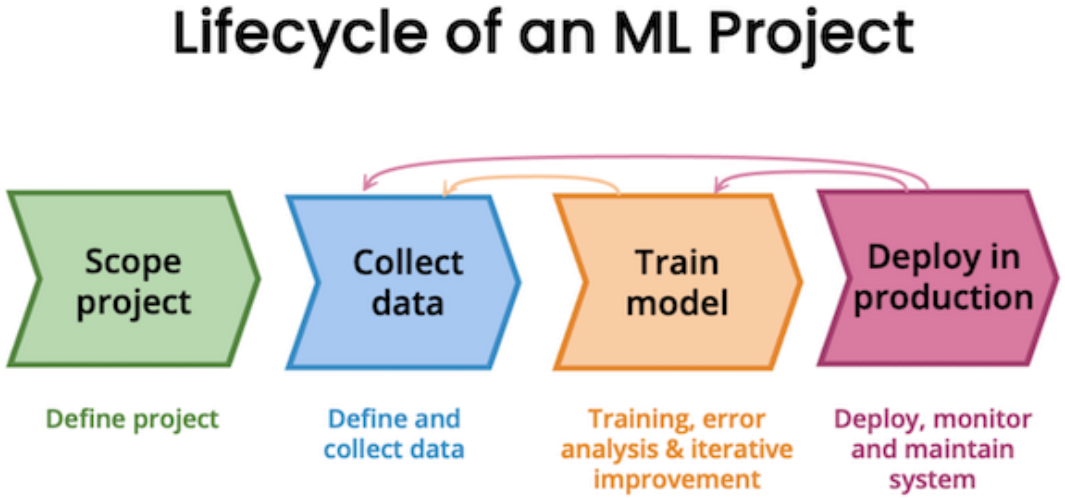
روش‌شناسی کاربردی

- بکارگیری موفقیت‌آمیز تکنیک‌های یادگیری عمیق به چیزی بیش از دانش کافی از الگوریتم‌ها و نحوه کار آنها نیاز دارد
- چگونه یک الگوریتم را برای یک کاربرد خاص انتخاب کنیم و چگونه می‌توان به بازخورد بدست آمده از آزمایش‌ها نظارت کرد و به آنها پاسخ داد؟
  - تصمیم برای جمع‌آوری داده‌های بیشتر
  - افزایش یا کاهش ظرفیت مدل
  - افزودن یا حذف منظّم‌سازی پارامترها
  - بهبود بهینه‌سازی یک مدل
  - اشکال‌زدایی نرم‌افزاری

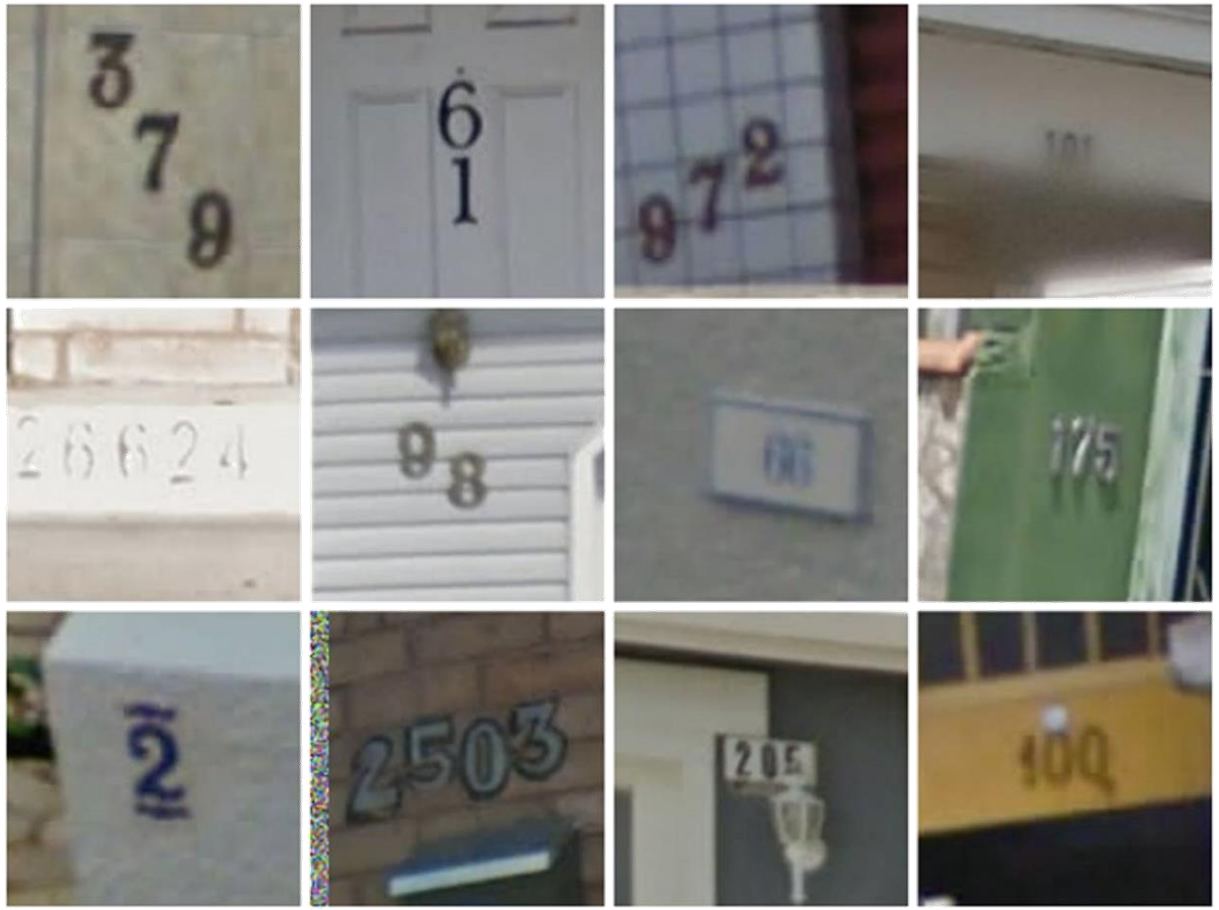


فرآیند طراحی

- اهداف خود را مشخص کنید
- در اسرع وقت یک مدل پایه end-to-end ایجاد کنید
- تشخیص دهید کدام بخش‌ها ضعیف‌تر از حد انتظار عمل می‌کنند
- به طور مکرر تغییرات تدریجی ایجاد کنید



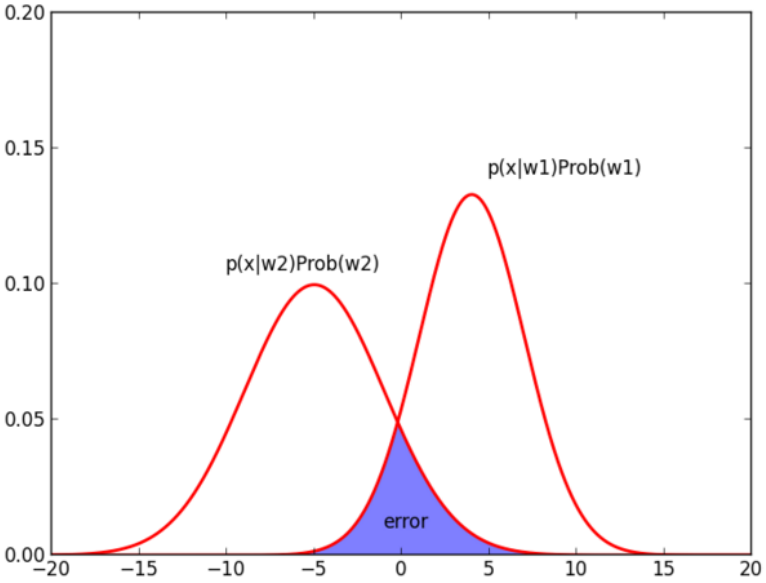
مثال: خواندن شماره پلاک ساختمان‌ها



- هدف افزودن ساختمان‌ها به نقشه گوگل است
- خودروهای Street View از ساختمان‌ها تصویر می‌گیرند و مختصات GPS مرتبط با هر تصویر را ثبت می‌کنند

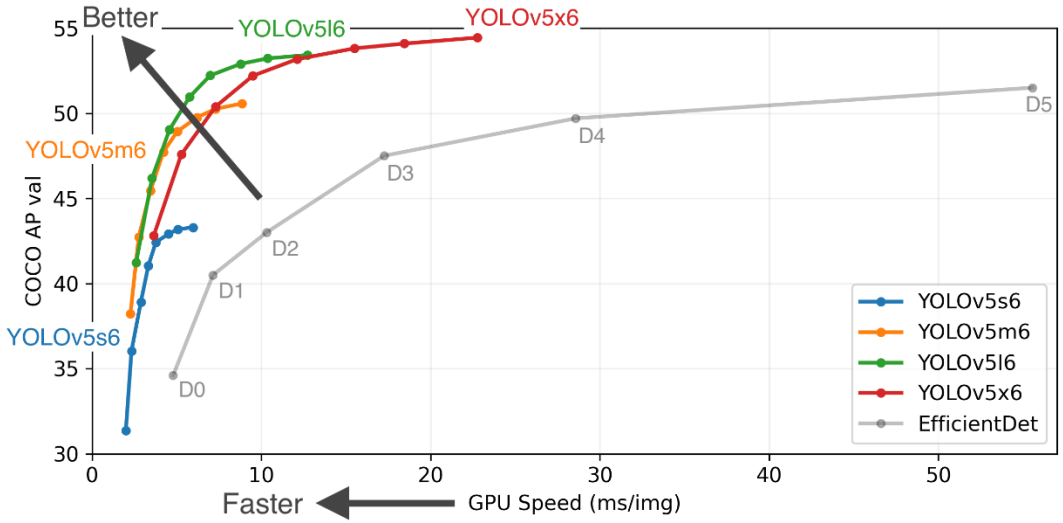
معیارهای ارزیابی عملکرد

- اهداف خود را مشخص کنید
  - از چه معیاری استفاده شود؟
  - چه سطحی از عملکرد مورد نظر است؟
- همه اقدامات آینده خود را با معیار خطا هدایت کنید
- هیچ برنامه‌ای به خطای صفر دست پیدا نمی‌کند!
- حجم داده‌های آموزشی به دلایل مختلفی محدودیت دارد
  - زمان، پول، دشواری



معیارهای ارزیابی عملکرد

- چگونه می‌توان سطح معقولی از عملکرد مورد انتظار را تعیین کرد؟



https://github.com/ultralytics/yolov5

- در محیط دانشگاهی
  - نرخ خطای قابل دستیابی بر اساس نتایج منتشر شده

- در محیط صنعتی
  - در مورد میزان حداکثر خطای ممکن برای ایمن، مقرون به صرفه یا جذاب بودن یک برنامه کاربردی برای مصرف‌کنندگان، ایده‌هایی وجود دارد