

رسالة محمد



مبانی بینایی کامپیوتر

مدرس: محمدرضا محمدی

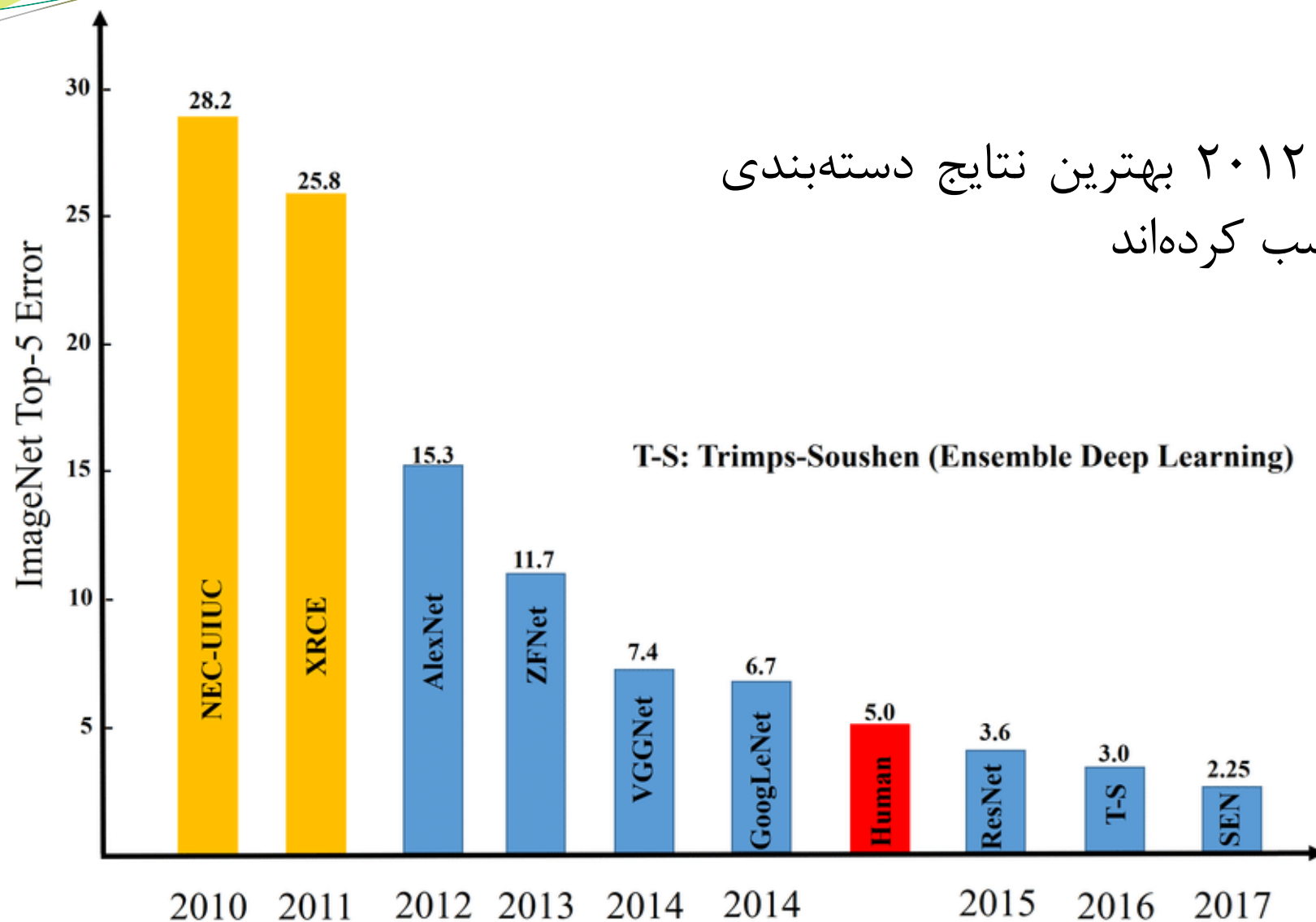
۱۳۹۹

معماری های CNN

CNN Architectures

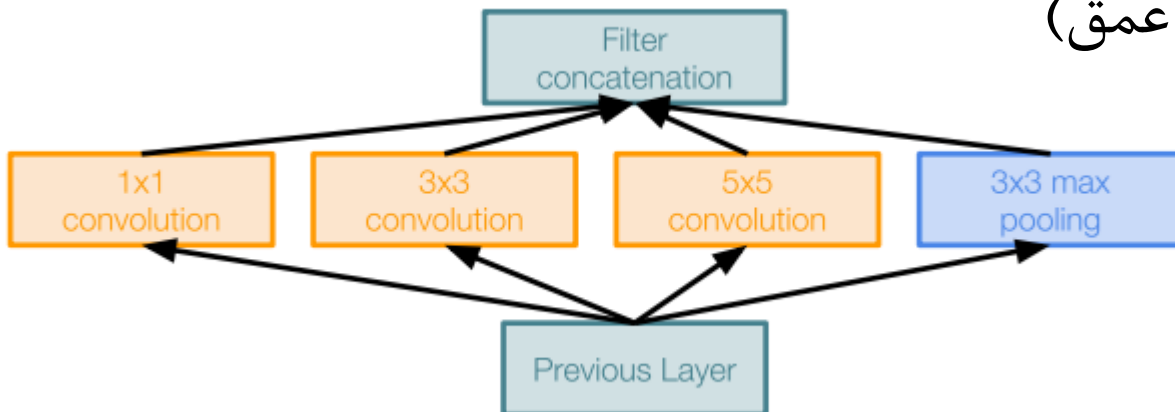
معماری‌های CNN

- معماری‌های مختلف CNN از سال ۲۰۱۲ بهترین نتایج دسته‌بندی تصویر در چالش ImageNet را کسب کرده‌اند



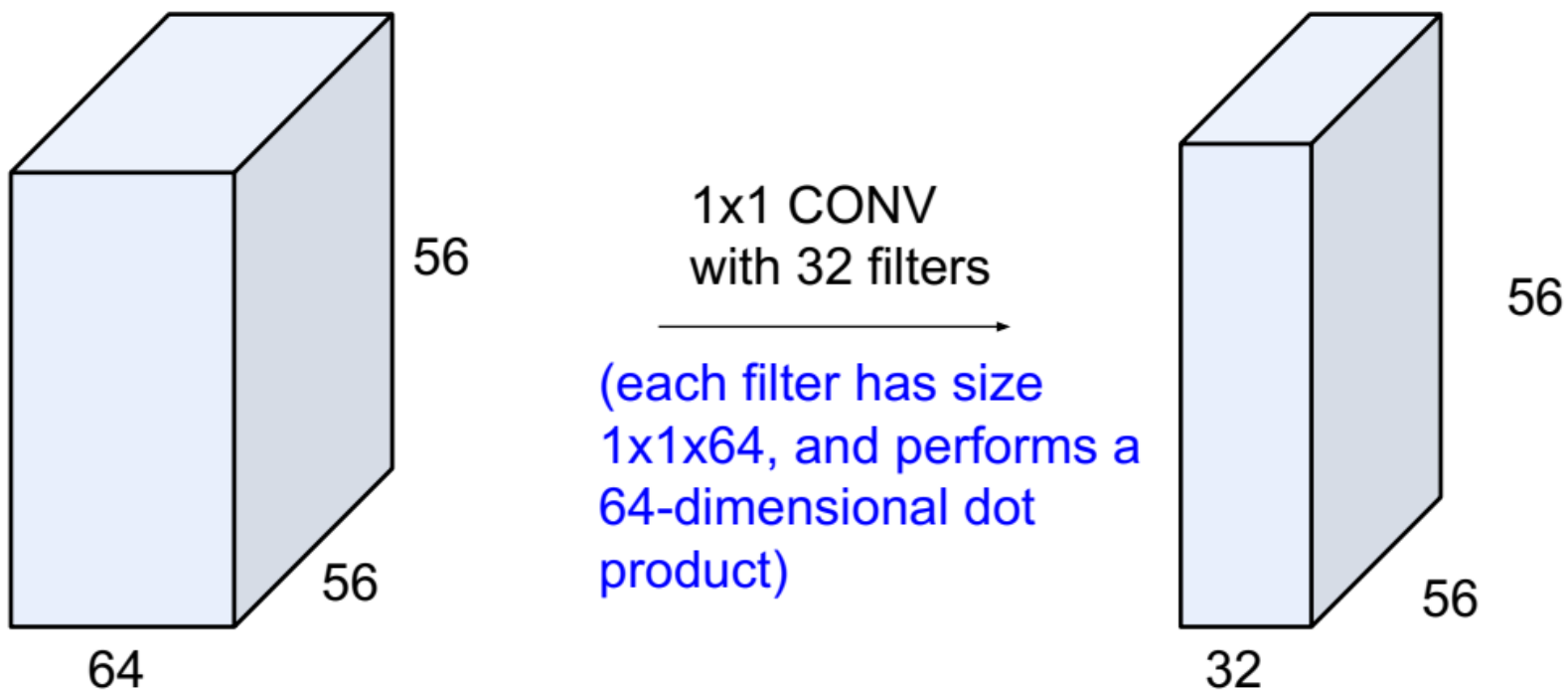
GoogLeNet

- شبکه GoogLeNet برنده مسابقه ILSVRC'14 با خطای ۶.۷٪ شد
- شبکه عمیق تر با پارامترهای کمتر
- فیلترهای هم‌عرض (موازی) تحت عنوان Inception Module معرفی شدند
 - کانولوشن‌های دارای ابعاد مختلف
 - عملیات Pooling
- سپس، خروجی تمام فیلترها به هم الحاق می‌شوند (در عمق)

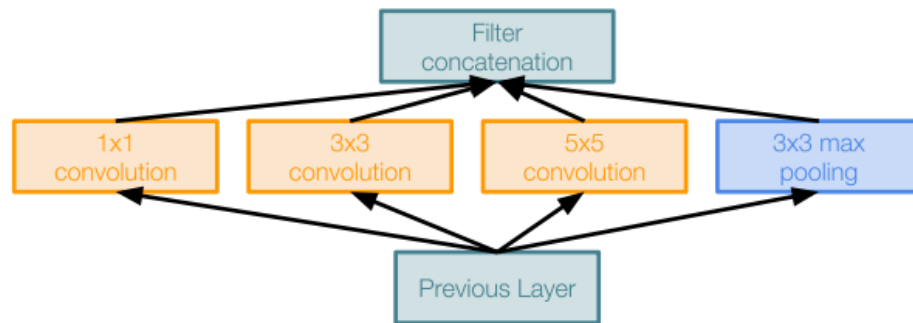


کانولوشن ۱×۱

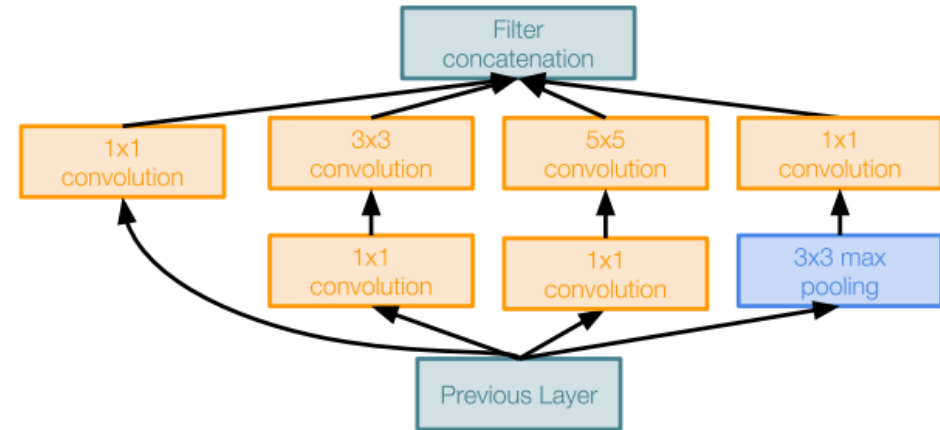
- ابعاد مکانی حفظ می شود و عمق کاهش می یابد



Inception Module



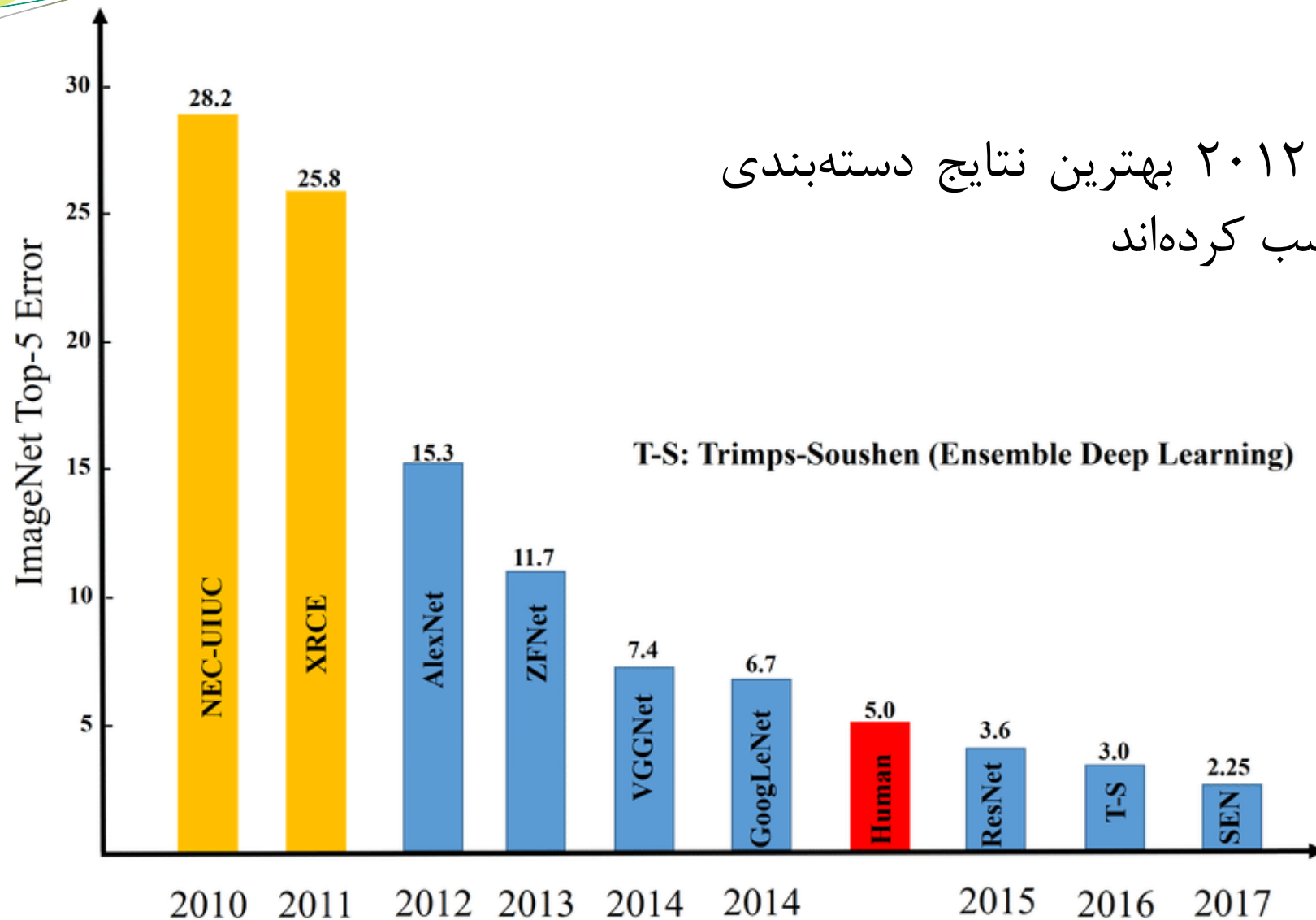
Naive Inception module



Inception module with dimension reduction

معماری‌های CNN

- معماری‌های مختلف CNN از سال ۲۰۱۲ بهترین نتایج دسته‌بندی تصویر در چالش ImageNet را کسب کرده‌اند



ResNet

Revolution of Depth

AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)

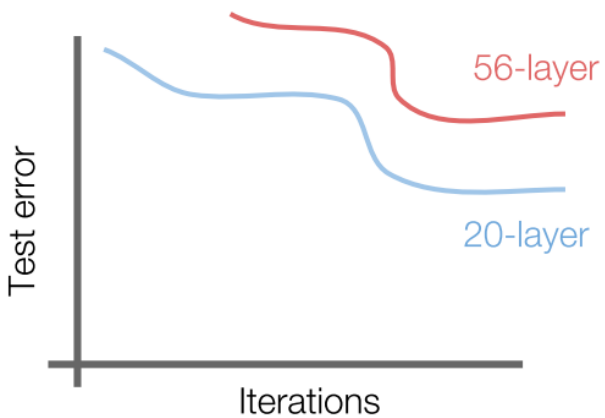
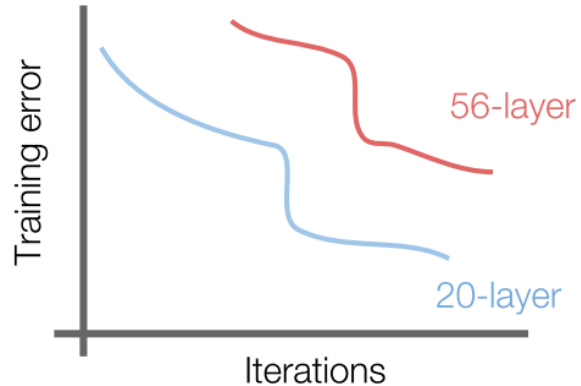


ResNet, 152 layers
(ILSVRC 2015)



- شبکه ResNet برنده مسابقه ILSVRC'15 با خطای ۳.۵۷٪ شد
- با ۱۵۲ لایه، انقلابی در عمق شبکه‌های کانولوشنی به وجود آورد

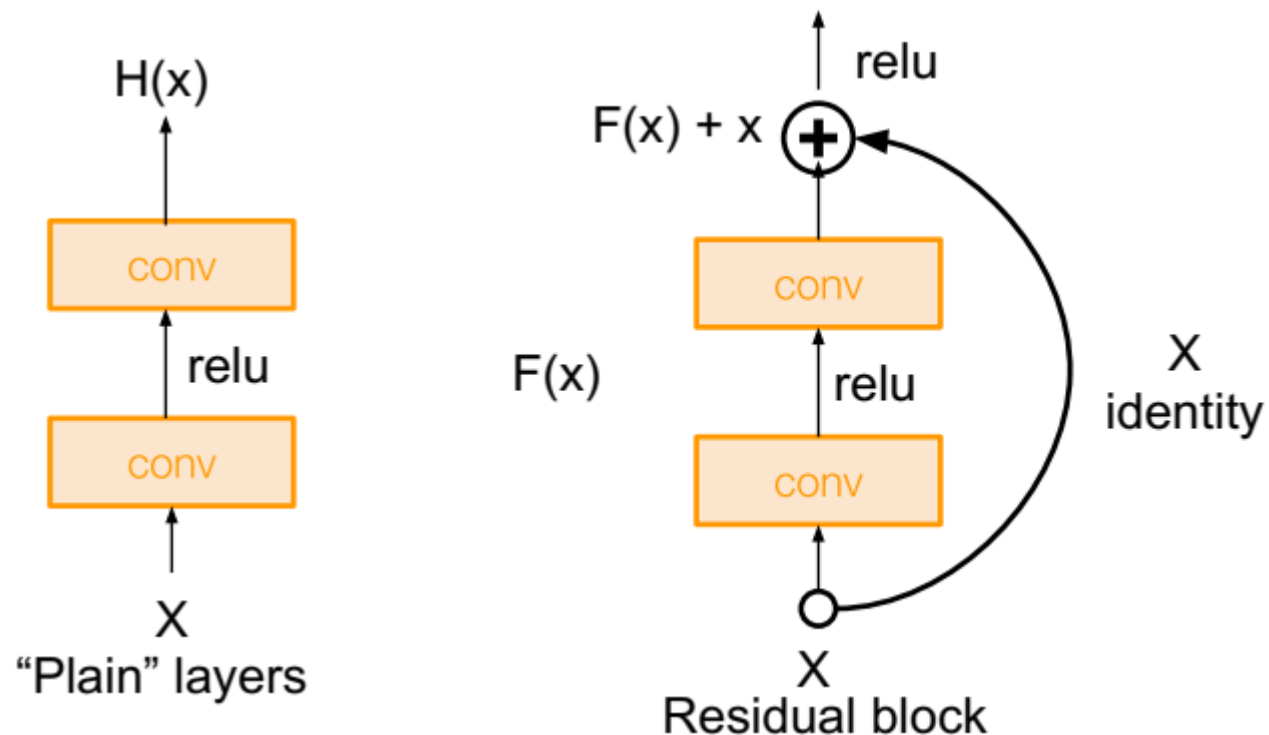
ResNet



- اگر تعداد لایه‌های کانولوشنی ساده را بسیار زیاد کنیم چه اتفاقی می‌افتد؟
- چرا شبکه عمیق‌تر هم در آموزش و هم در آزمون عملکرد ضعیف‌تری دارد؟
 - البته مشکل از overfitting نیست!
- فرضیه: مشکل در مسئله بهینه‌سازی است
 - بهینه‌سازی مدل‌های عمیق‌تر دشوارتر است
- عملکرد مدل‌های عمیق‌تر باید حداقل به خوبی مدل‌های با عمق کمتر باشد
 - می‌توان وزن‌های مدل کم‌عمق را به لایه‌های نخست شبکه عمیق کپی کرد و لایه‌های اضافی را به گونه‌ای تنظیم کرد که نگاشت همانی را انجام دهند
- ایده ResNet آن است که لایه‌های شبکه بجای آموختن نگاشت مطلوب، باقی‌مانده آن را یاد بگیرند

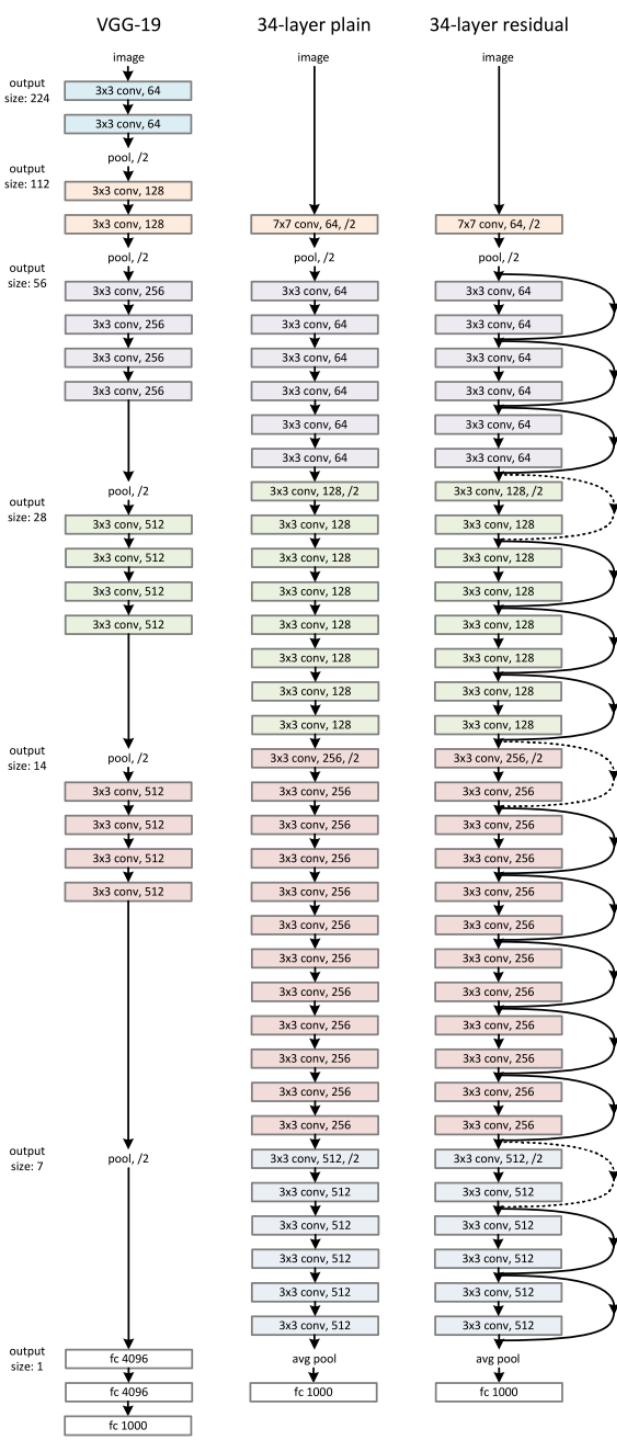
ResNet

- لایه‌ها باید $F(x) = H(x) - x$ را بجای $H(x)$ بیاموزند



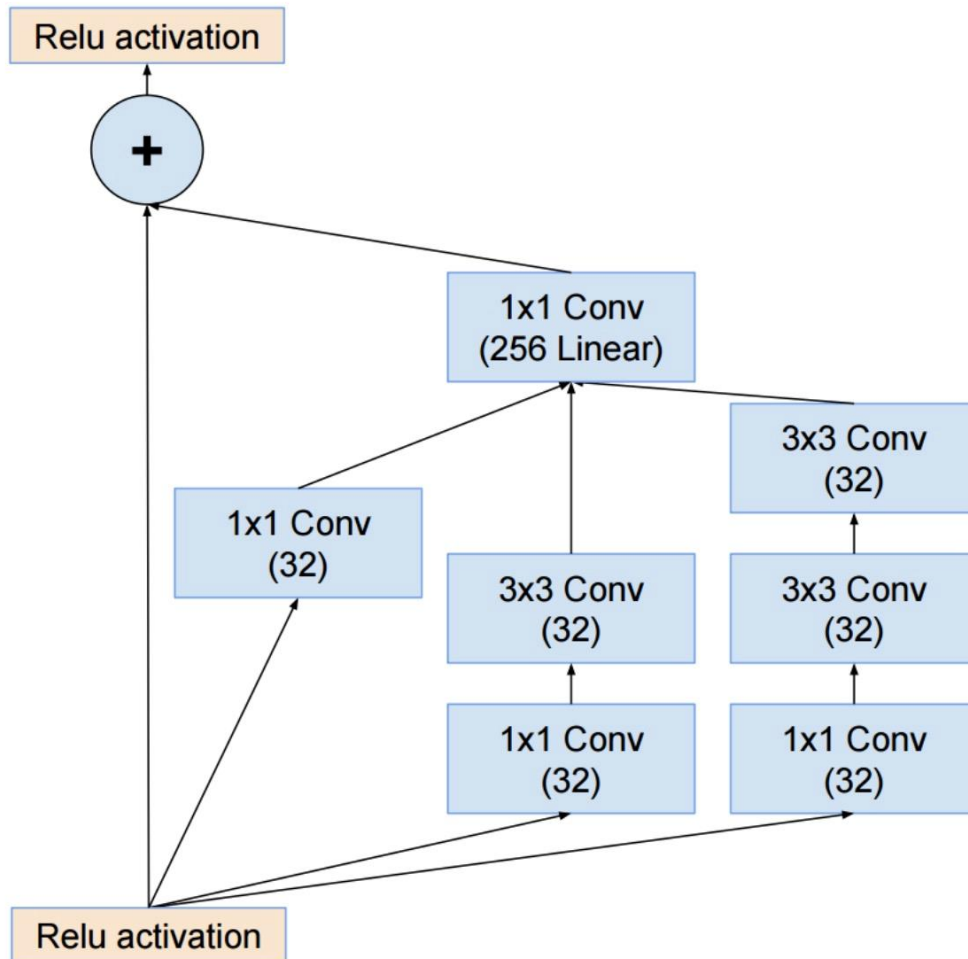
ResNet

- از تعداد زیادی بلوک باقی مانده تشکیل شده است
- هر بلوک باقی مانده دارای ۲ لایه کانولوشنی 3×3 است
- به طور دوره‌ای، تعداد فیلترها ۲ برابر شده و رزولوشن مکانی نصف می‌شود
- در ابتدا دارای یک لایه کانولوشنی است
- پس از آخرین بلوک باقی مانده، ابعاد داده‌ها با استفاده از Average Pooling کاهش می‌یابد و یک لایه FC برای دسته‌بندی استفاده می‌شود
- برای مسئله ImageNet عمق‌های مختلف شبکه شامل ۳۴، ۵۰، ۱۰۱ و ۱۵۲ استفاده شده‌اند

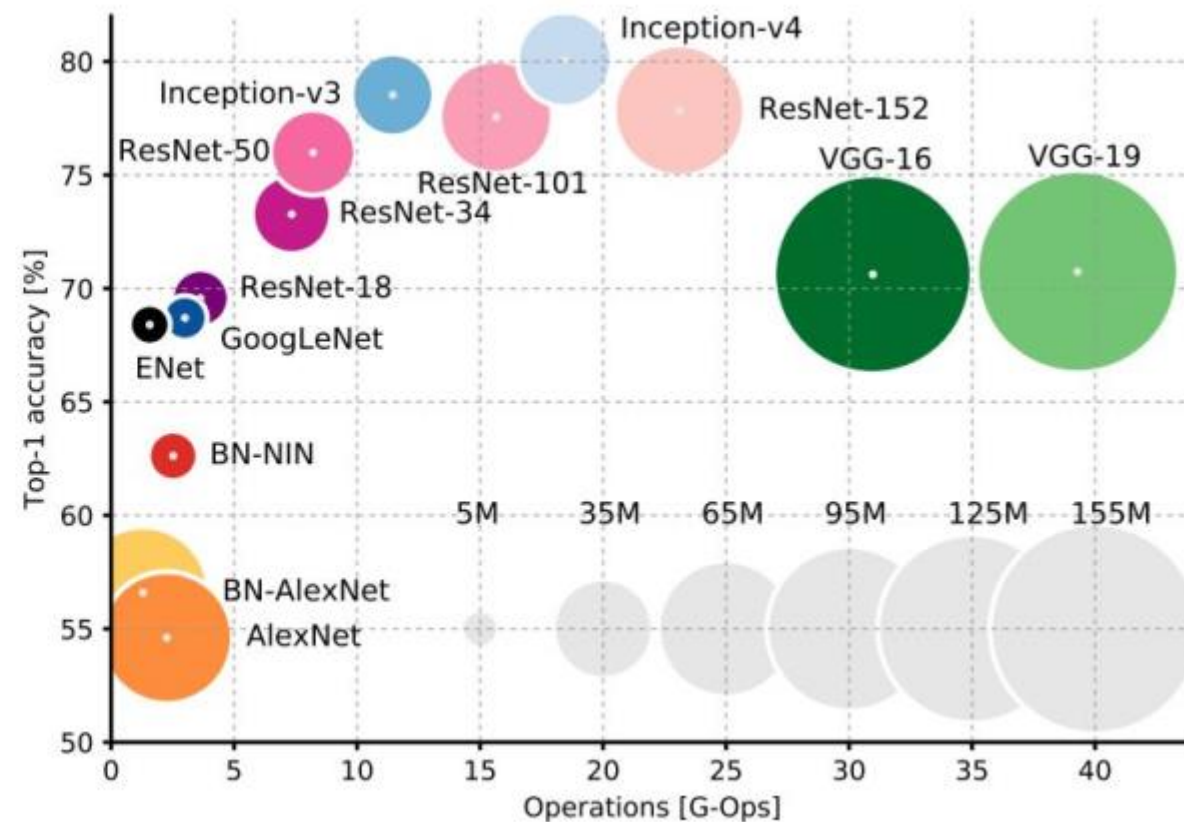
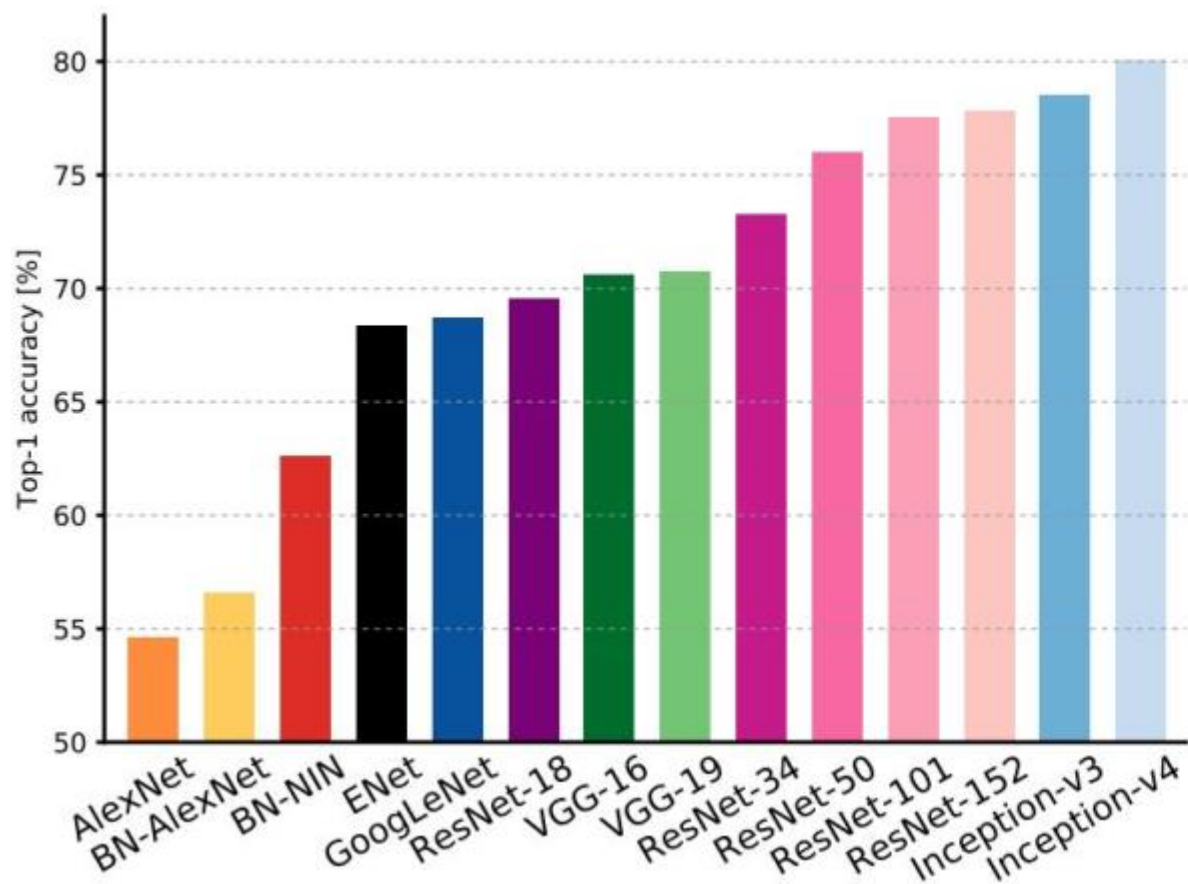


Inception V4

- ورودی همانی به ماژول Inception افزوده شده است



مقایسه

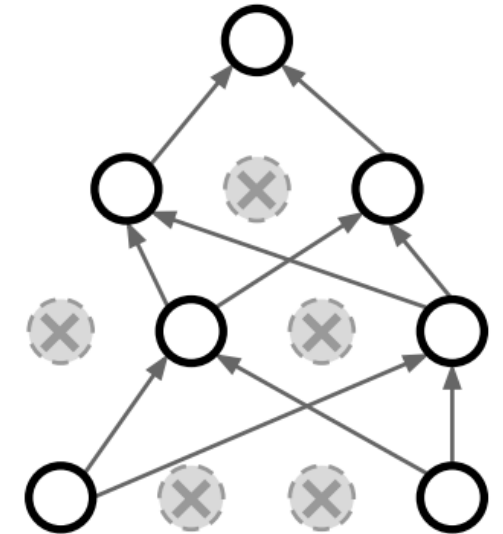
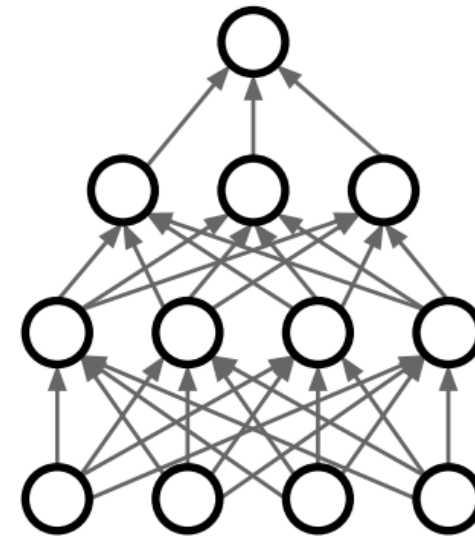
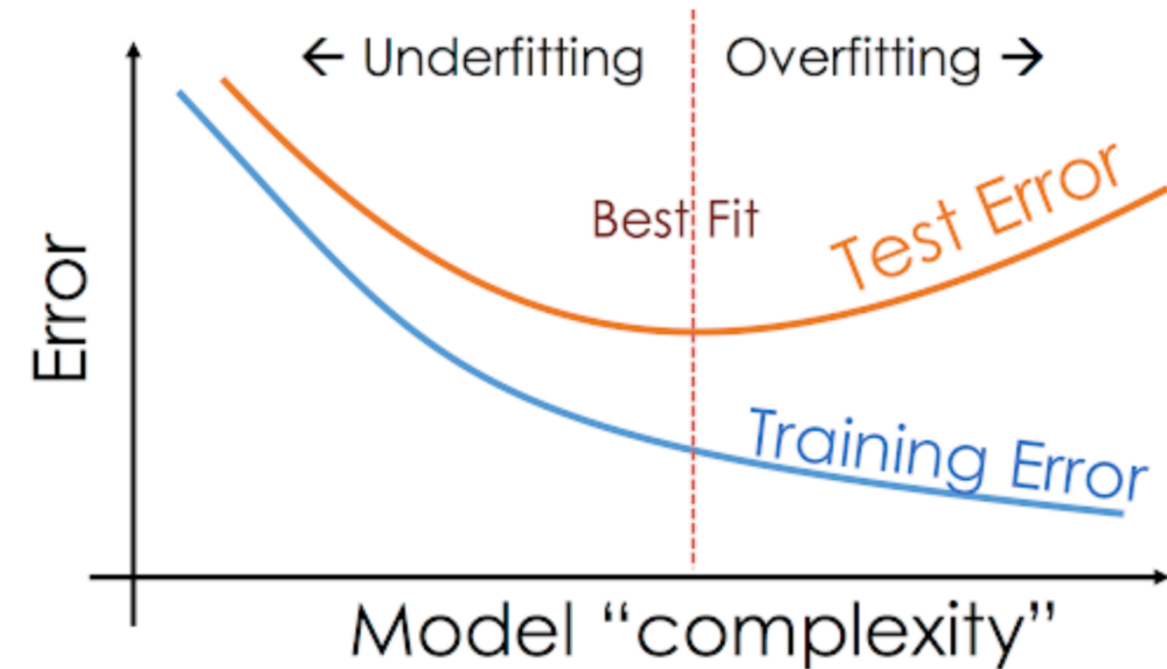


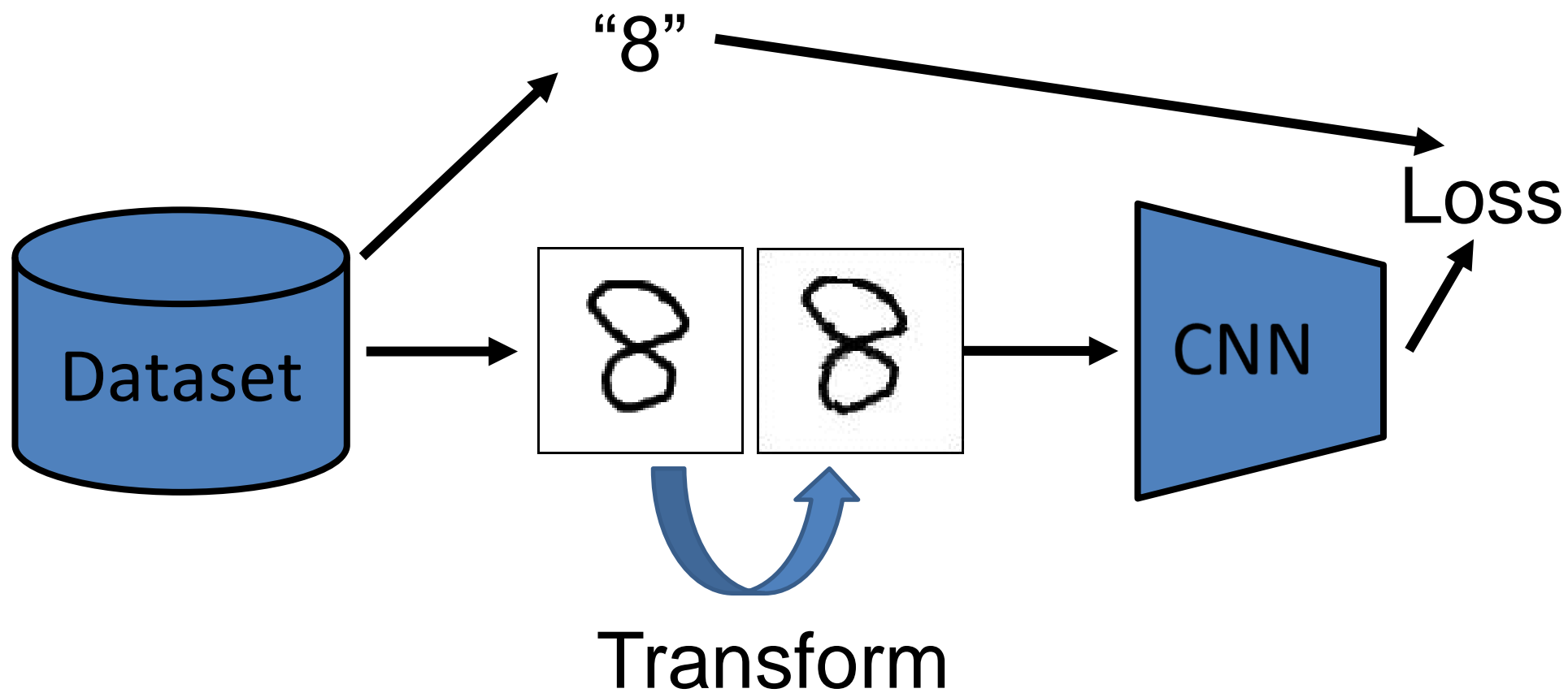
داده‌افزایی

Data Augmentation

Overfitting vs Underfitting

- برای جلوگیری از overfitting و استفاده از مزایای شبکه‌های دارای توانایی یادگیری بیشتر، می‌توان مسئله‌ای که قرار است توسط شبکه حل شود را پیچیده‌تر کرد
- استفاده از dropout نمونه‌ای از این موارد است





داده‌افزایی: Flip



6

د



e

9