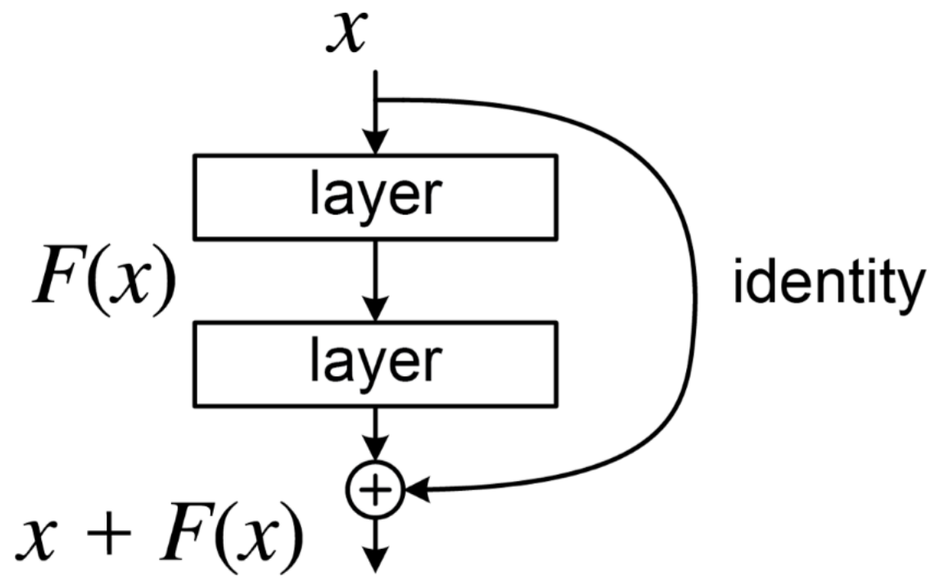
Day-20, Oct-22, 2024

CNN continuing — Resnet

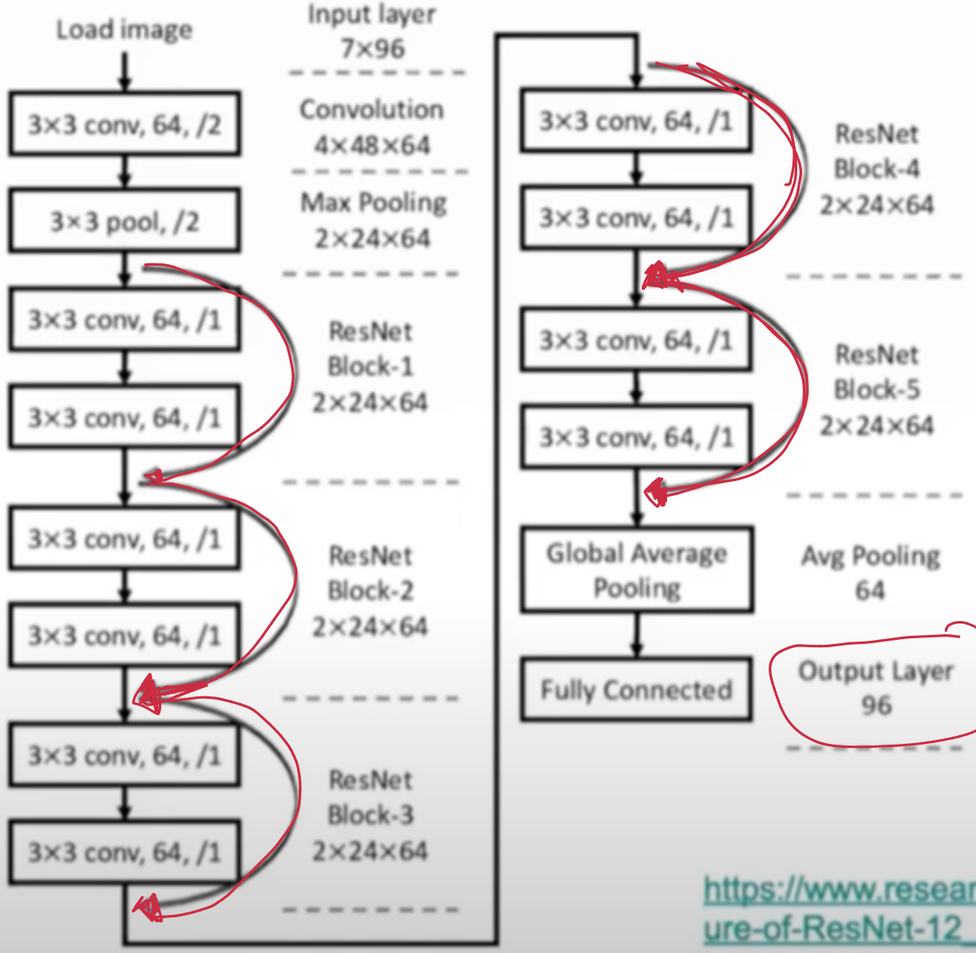
Residual Neural Networks (Resnet)

introduced novel concept 'skip connections'. It allow information to flow directly from earlier layers to later ones, bypassing intermediate layers. Direct path, Residual Blocks, Identity function, and solves vanishing gradient problems.



Resnet

3×3 → Kernal Size
64 → filters
/2 → strides



Skip Connections.

Convolution
4×48×64
4 → rows
48 → columns
64 → Filters
→ 64 Feature Maps

[Identity function As shortcut]

I/p 7×96
O/p:

→ O/p - 96 classification.

https://www.researchgate.net/figure/The-structure-of-ResNet-12_fig1_329954455

nepalschool.naamii.org.np

Identity function — input x to the function f(x) → x 'output'

Resnet → SKip Connections, layers (shortcut-connections)

THIRD NEPAL WINTER SCHOOL in AI, 2021
NepAl Applied Mathematics and Informatics Institute for research

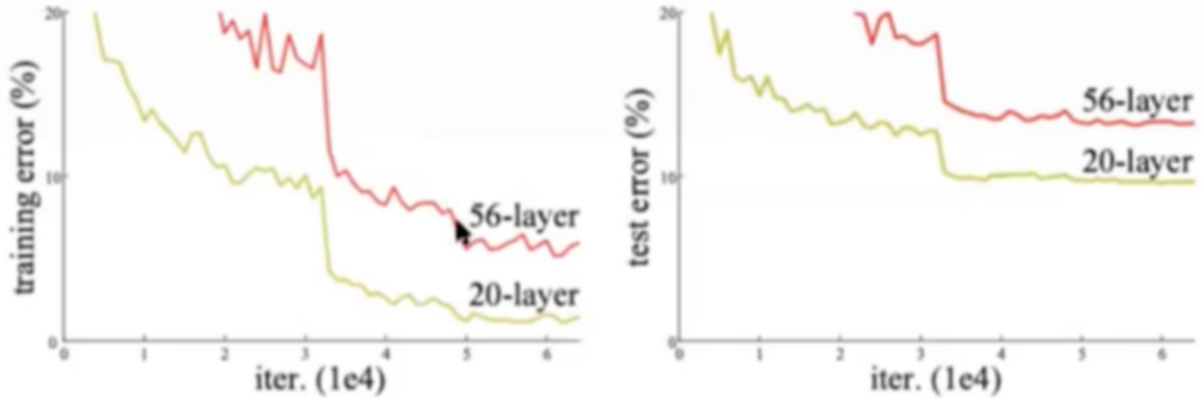


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

nepalschool.naamii.org.np

As we add more layers to the deep Neural Networks then the problems or the AND becomes unstabilize so we can use Resnet.

Figure 1 (18-layers No Resnet) (34-layers Resnet)

Fig 1.
Resnet-18 (18 layers)

Fig 2.
Resnet-34 (layers 34)

Resnet

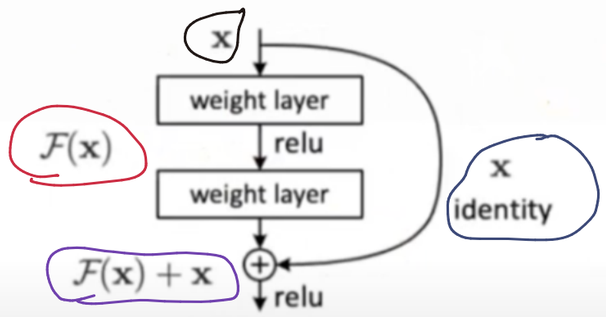


Figure 2. Residual learning: a building block.

- Assuming that the original desired mapping is H(x):
  - Before: fit F(x) = H(x) - x
  - Resnet: fit H(x) = F(x) + x
- The hypothesis was that it's easier to fit the residual mapping than the original.
- In extreme case, easier to map to 0 than map to 1.

Fig-2.
Resnet-34 perform better than Resnet-18 due to Resnet Applied.

Red → 34
Blue → 18

Fig 1.
Resnet-34 has higher error than Resnet-18

plots Comparing the training and validation error curves for different ResNet Architecture.

Resnet

Fig 1
Without Resnet X

Fig 2.
With Resnet ✓

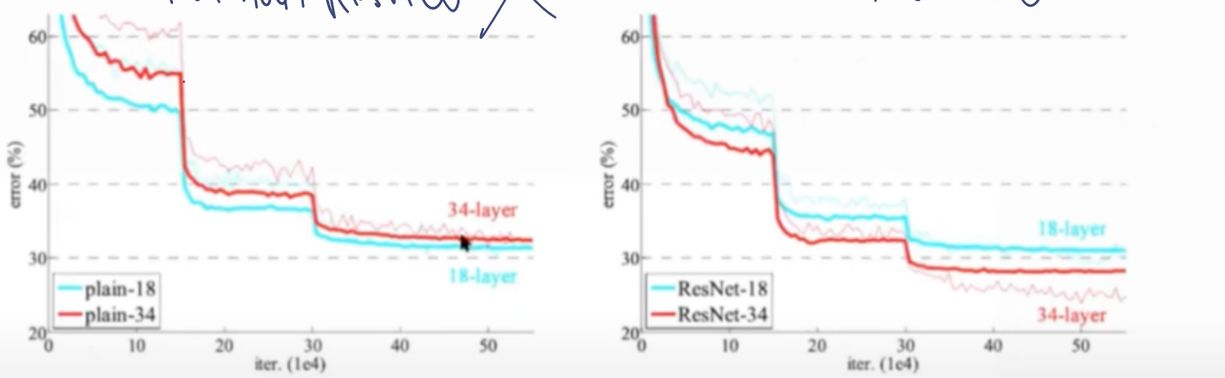


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

GoogleNet

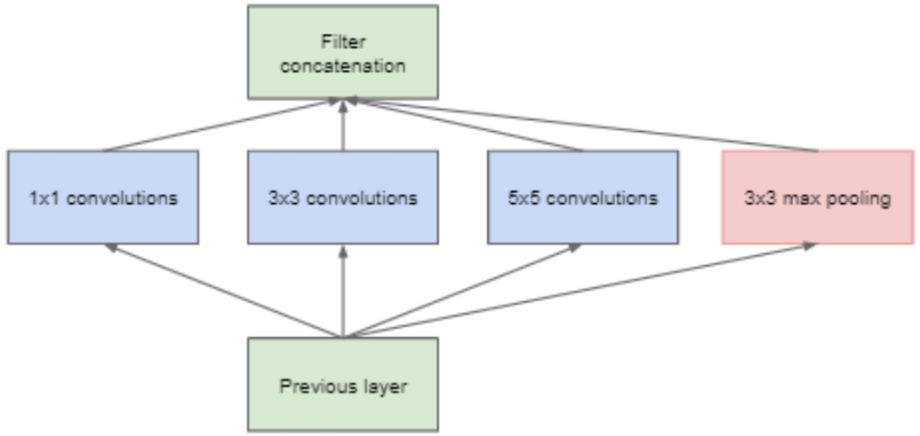
- Does dimensionality reduction on each layer and does convolutions at different resolutions to build sparse but more diverse/representative features.
- "visual information should be processed at various scales and then aggregated so that the next stage can abstract features from different scales simultaneously"

Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

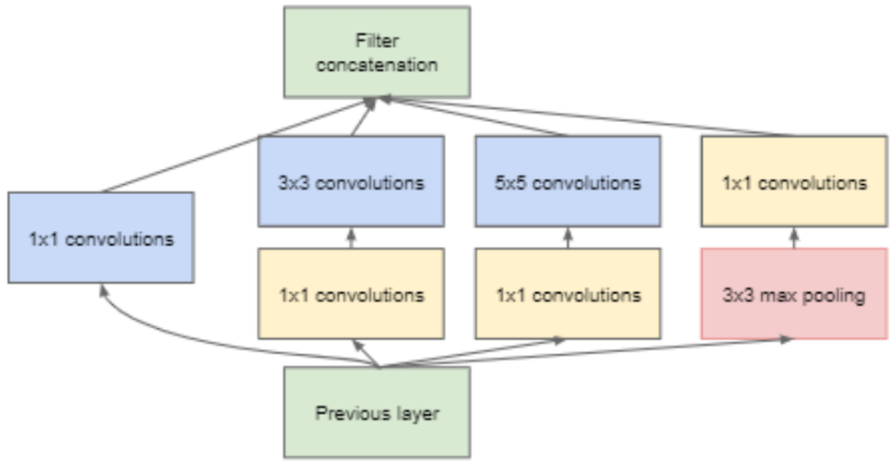
THIRD NEPAL WINTER SCHOOL in AI, 2021
NepAl Applied Mathematics and Informatics Institute for research

nepalschool.naamii.org.np

GoogleNet (Inception Module) allows the network to choose between several Convolutional filter sizes in each layer that is beneficial for capturing both local and global features efficiently.



(a) Inception module, naïve version



(b) Inception module with dimension reductions

[1x1 is multiplied by constant]

Auxiliary Classifier:

→ Tackle gradient vanishing problem

→ provide additional gradient flow to the previous layer

→ Incorporated at the intermediate layers.

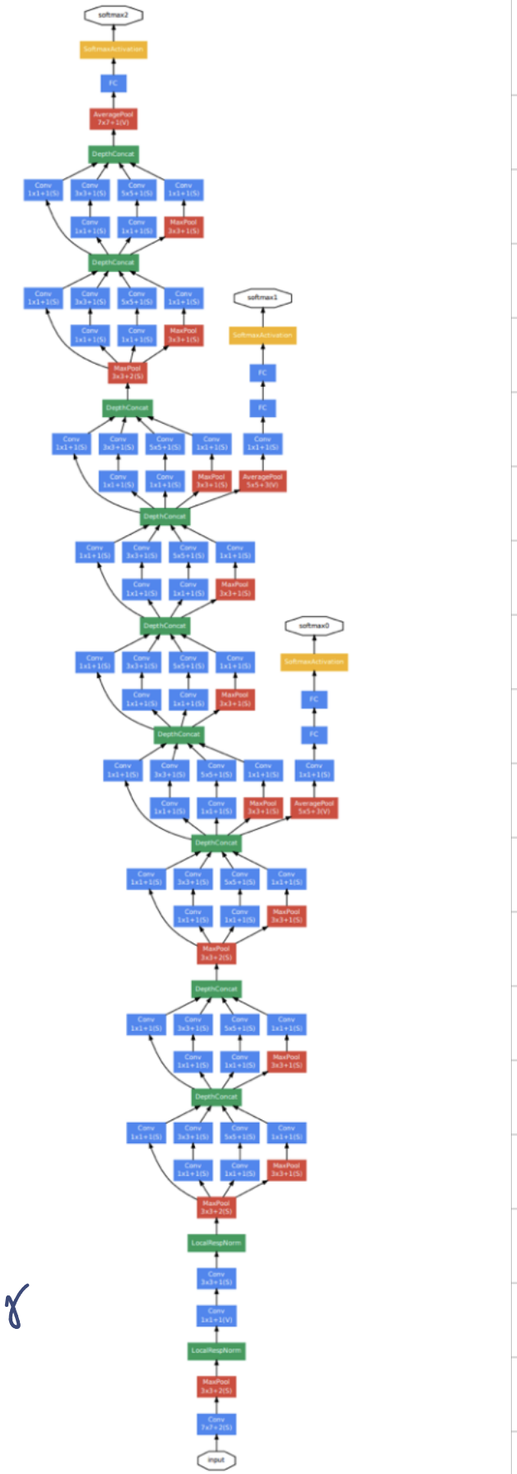


Figure 3: GoogLeNet network with all the bells and whistles

GoogleNet

- Uses 7X7 convolutions after the input - which is rather unusual.
- Auxiliary classifiers:
  - The model had two additional outputs, trying to predict the same thing as the primary output.
  - These are added in the middle part of the network - in order to increase the discriminative power of the middle layers.
  - The losses are added to overall loss, but multiplied by 0.3
  - This helped in training the model better and not having bottlenecks in the middle layers.

THIRD NEPAL WINTER SCHOOL in AI, 2021
NepAl Applied Mathematics and Informatics Institute for research

See the Report of 2014 GoogleNet.

GoogleNet on ImageNet

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

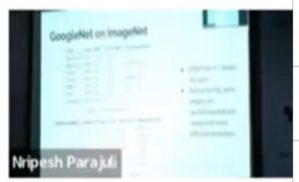
Table 2: Classification performance

Number of models	Number of Crops	Cost	Top-5 error	compared to base
1	1	1	10.07%	base
1	10	10	9.15%	-0.92%
1	144	144	7.89%	-2.18%
7	1	7	8.09%	-1.98%
7	10	70	7.62%	-2.45%
7	144	1008	6.67%	-3.45%

Table 3: GoogLeNet classification performance break down

THIRD NEPAL WINTER SCHOOL in AI, 2021
NepAl Applied Mathematics and Informatics Institute for research

- Ensemble of 7 models are used.
- During testing, same images are resized/rescaled and cropped at many different resolutions.

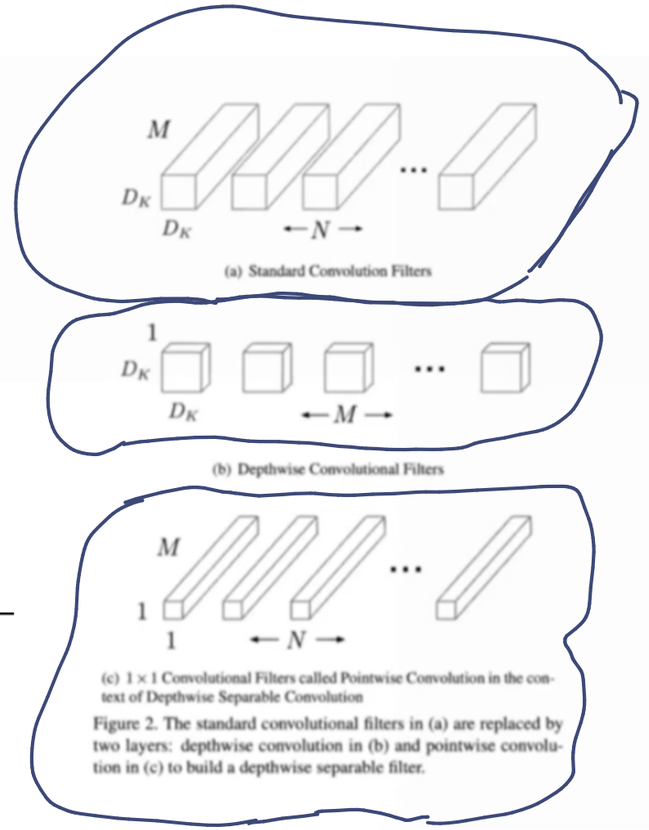


nepalschool.naamii.org.np

Mobilenet are efficient & lightweight for mobile devices.

Mobilenet

- Uses separable depthwise convolution where regular convolution - it is separated into depthwise convolution (3 X 3), followed by (1 X 1) convolution.
- This reduces number of parameters by about 8-9 times.



(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Figure 2. The standard convolutional filters in (a) are replaced by two layers: depthwise convolution in (b) and pointwise convolution in (c) to build a depthwise separable filter.

Mobilenet

Table 8. MobileNet Comparison to Popular Models

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138

Table 9. Smaller MobileNet Comparison to Popular Models

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
0.50 MobileNet-160	60.2%	76	1.32
Squeezenet	57.5%	1700	1.25
AlexNet	57.2%	720	60

- For comparison, resnet-152 had 79% accuracy (top-1 error). But is so much bigger and slower.

Mobilenet

- The model allows for reducing the size / latency further by using less channels (controlled by alpha), and by reducing input image resolution (controlled by ro).
- Small reduction in number of channels and resolution, still gives good performance.

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
Conv MobileNet	71.7%	4866	29.3
MobileNet	70.6%	569	4.2

Table 5. Narrow vs Shallow MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
0.75 MobileNet	68.4%	325	2.6
Shallow MobileNet	65.3%	307	2.9

Table 6. MobileNet Width Multiplier

Width Multiplier	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
0.75 MobileNet-224	68.4%	325	2.6
0.5 MobileNet-224	63.7%	149	1.3
0.25 MobileNet-224	50.6%	41	0.5

Table 7. MobileNet Resolution

Resolution	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
1.0 MobileNet-192	69.1%	418	4.2
1.0 MobileNet-160	67.2%	290	4.2
1.0 MobileNet-128	64.4%	186	4.2

THIRD NEPAL WINTER SCHOOL in AI, 2021
NepAl Applied Mathematics and Informatics Institute for research

nepalschool.naamii.org.np

→ MobileNet are especially made for mobile and embedded visual applications.

→ Resnet can handle deep neural network solving vanishing gradient problem, skip connectism.

Mobilenet	Resnet	Googlenet
→ For mobile, embedded visual system & less Computational devices	→ Residual Blocks (Skip Connections)	→ Inception Modules (1x1, 3x3, 5x5)
→ Depthwise Separable Convolutions, faster (~ 4.2 million)	→ Can be very deep accurate models for classification but costly (~ 25 million)	→ features various scales with lower parameter (~ 5 million)
→ May Sacrifice too much accuracy.	→ Deep learning visual task	→ General vision task

From Textbook: # Knowledge Representation: Unification.

Unification:

Unification is a process in FOPL that aims to make two logical expression identical by finding a substitution (set of variable replacements).

Example:

→ Expressions: $p(x, a)$ and $p(b, y)$

→ Unification: By substituting $x = b$ & $y = a$ these two expressions become identical

$p(b, a)$ and $p(b, a)$ ← Substituting.

Lifting:

Lifting refers to extending the inference process from propositional logic to first-order-logic, variables are involved, predicates involved.

It is the process of transforming a ground clause (a clause without variables) into the general clause (a clause with variables).

$$\neg P(a) \vee Q(b) \longrightarrow \neg P(x) \vee Q(y)$$

Ground Clause → General Clause

Generalizing Resolution, Avoiding Grounded Inference, Lifting Inference to FOPL

and Role of unification in lifting are some points to be considered.

Example: Suppose you have two clauses:

$P(x) \longrightarrow$ where x is variables

$\neg P(a) \longrightarrow$ where a is Constant

Before applying Inference (resolution), we must "lift" the process by unifying $x = a$, making the clauses Compatible.

① $P(x) \vee Q(x)$ ② $\neg P(a)$ ③ $P(a) \vee P(a)$

(1) Initial clauses:

(1) clause 1: $p(x) \vee Q(x)$

(2) clause 2: $\neg p(a)$

(2) Unification: $x = a$.

clause 1: $p(a) \vee Q(a)$

clause 2: $\neg p(a)$

(3) Applying Resolution: clause 1 : $p(a) \vee Q(a)$

clause 2: $\neg p(a)$

Resolution step:

Complementary literals: $\neg P(a)$

$$\frac{(P(a) \vee Q(a)) \qquad \neg P(a)}{Q(a)}$$

(8) Result → $Q(a)$. → Holds true.

Machine vision

→ Application-Specific, real-time Processing, decision-making, Automation

→ Heavy Realiance on Specialized Hardware

Computer vision

→ General-purpose, understanding & Interpreting visual data

→ ML, DL Algorithms, Object Recognition