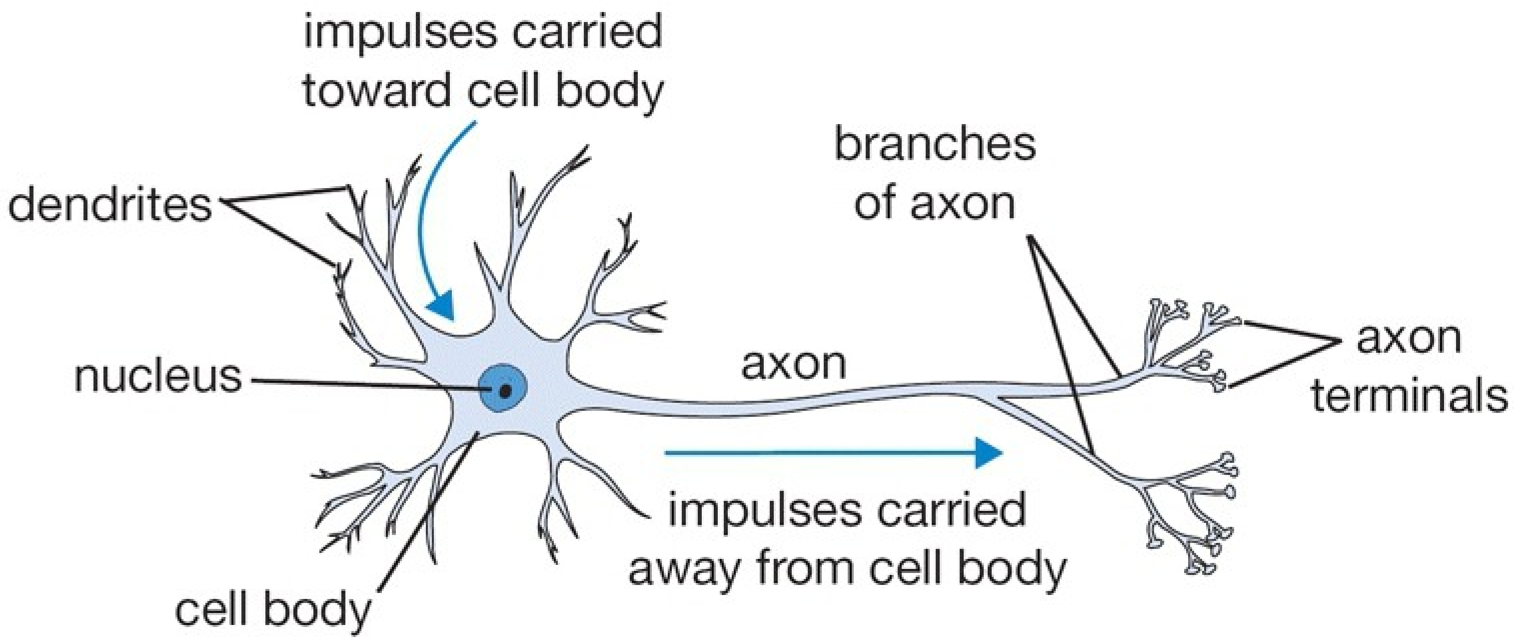
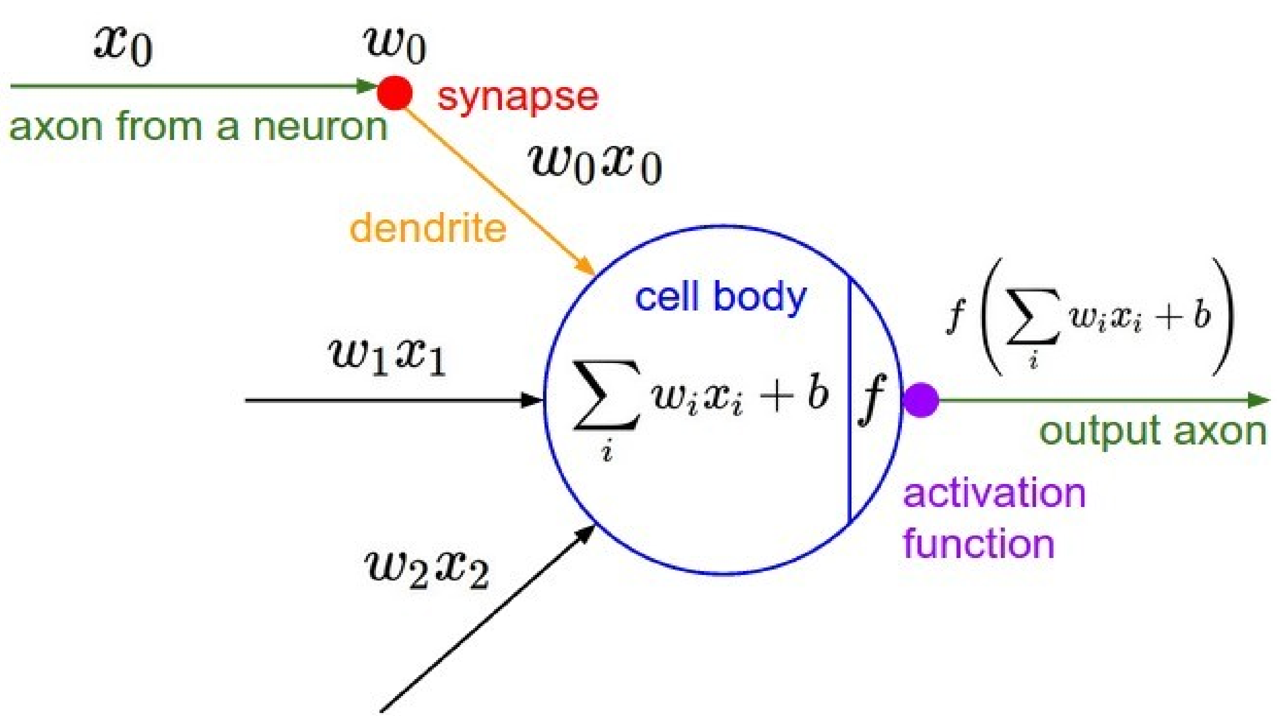
Deep neural networks in a nutshell

Many inventions were inspired by the Nature


AIR CHINA
B-2947


impulses carried
toward cell body
dendrites
branches
of axon
axon
terminals
nucleus
axon
impulses carried
away from cell body
cell body

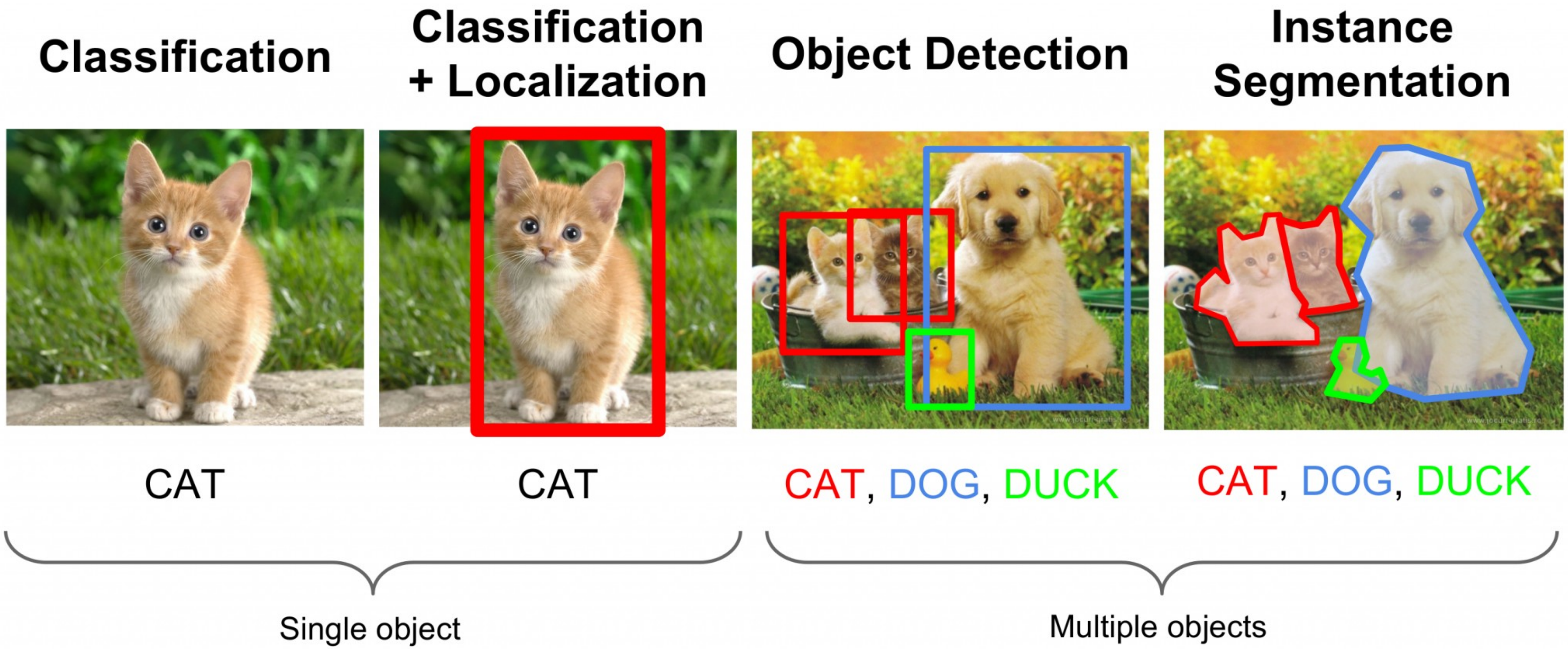

x_0
w_0
synapse
axon from a neuron
$w_0 x_0$
dendrite
cell body
$f\left(\sum_i w_i x_i + b\right)$
$w_1 x_1$
$\sum_i w_i x_i + b$
f
output axon
activation
function
$w_2 x_2$

Incredibly poor analogy from biological point of view

It doesn't matter how you come up with the idea, but its utility is the only thing that matters.

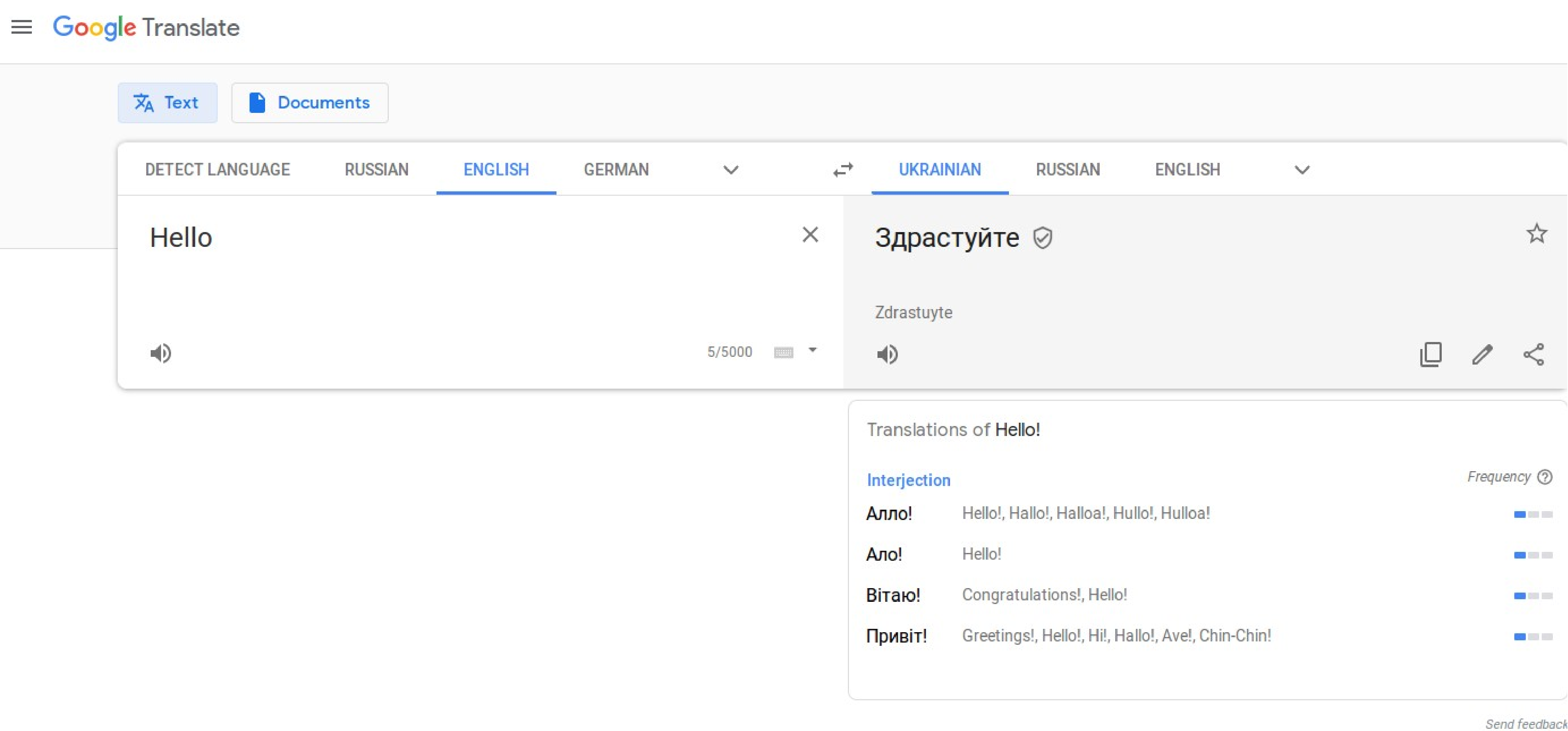
NN Tasks and Utility

- Classification
- Classification with missing inputs
- Segmentation
- Regression (predict a numerical value given some input)
- Transcription (observe a relatively unstructured representation of some kind of data and transcribe it into discrete, textual form)
- Machine translation
- Structured output (e.g. sentence to its grammar tree)
- Anomaly detection
- Synthesis and sampling
- Imputation of missing values
- Denoising
- Density estimation or probability mass function estimation
- And many-many more.....


Classification
Classification + Localization
Object Detection
Instance Segmentation
CAT
CAT
CAT, DOG, DUCK
CAT, DOG, DUCK
Single object
Multiple objects

Classification is simple? Think once more.

Google uses NN for translation



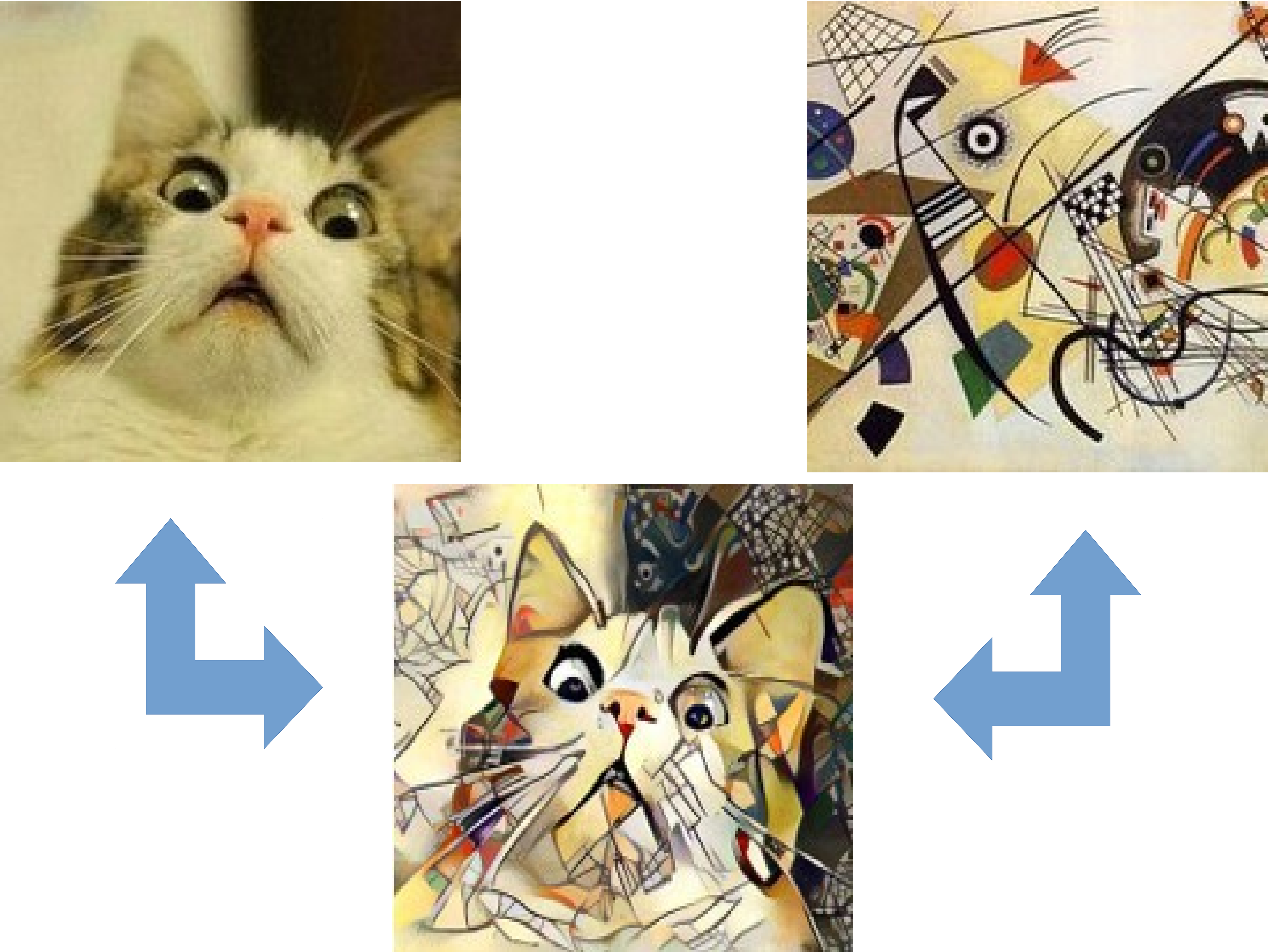

self, etc.
dynamic
ground
road
sidewalk
parking
rail track
building
wall
fence
guard rail
bridge
tunnel
pole
polegroup
traffic light
traffic sign
vegetation
terrain
sky
person
rider
car
truck
bus
caravan
trailer
train
motorcycle
bicycle

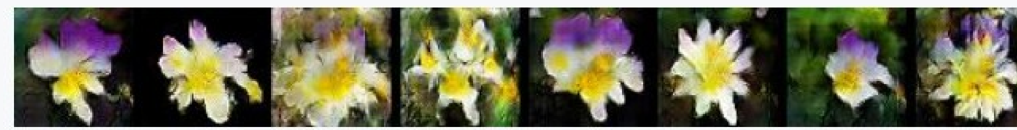
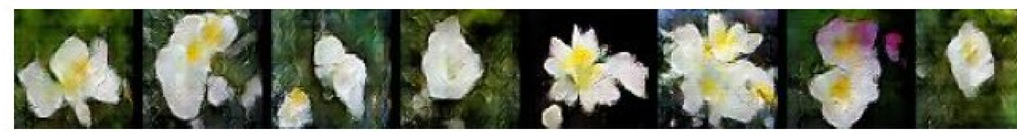
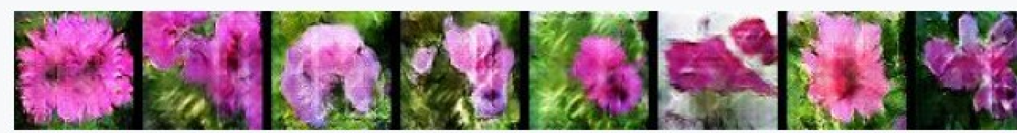
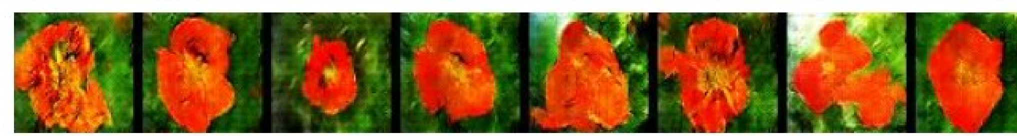
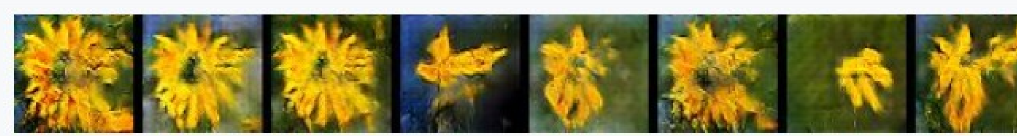

LEFT REARWARD VEHICLE CAMERA
MEDIUM RANGE VEHICLE CAMERA
RIGHT REARWARD VEHICLE CAMERA
MOTION FLOW
LANE LINES
LANE LINES
ROAD FLOW
IN-PATH OBJECTS
ROAD LIGHTS
OBJECTS
ROAD SIGNS

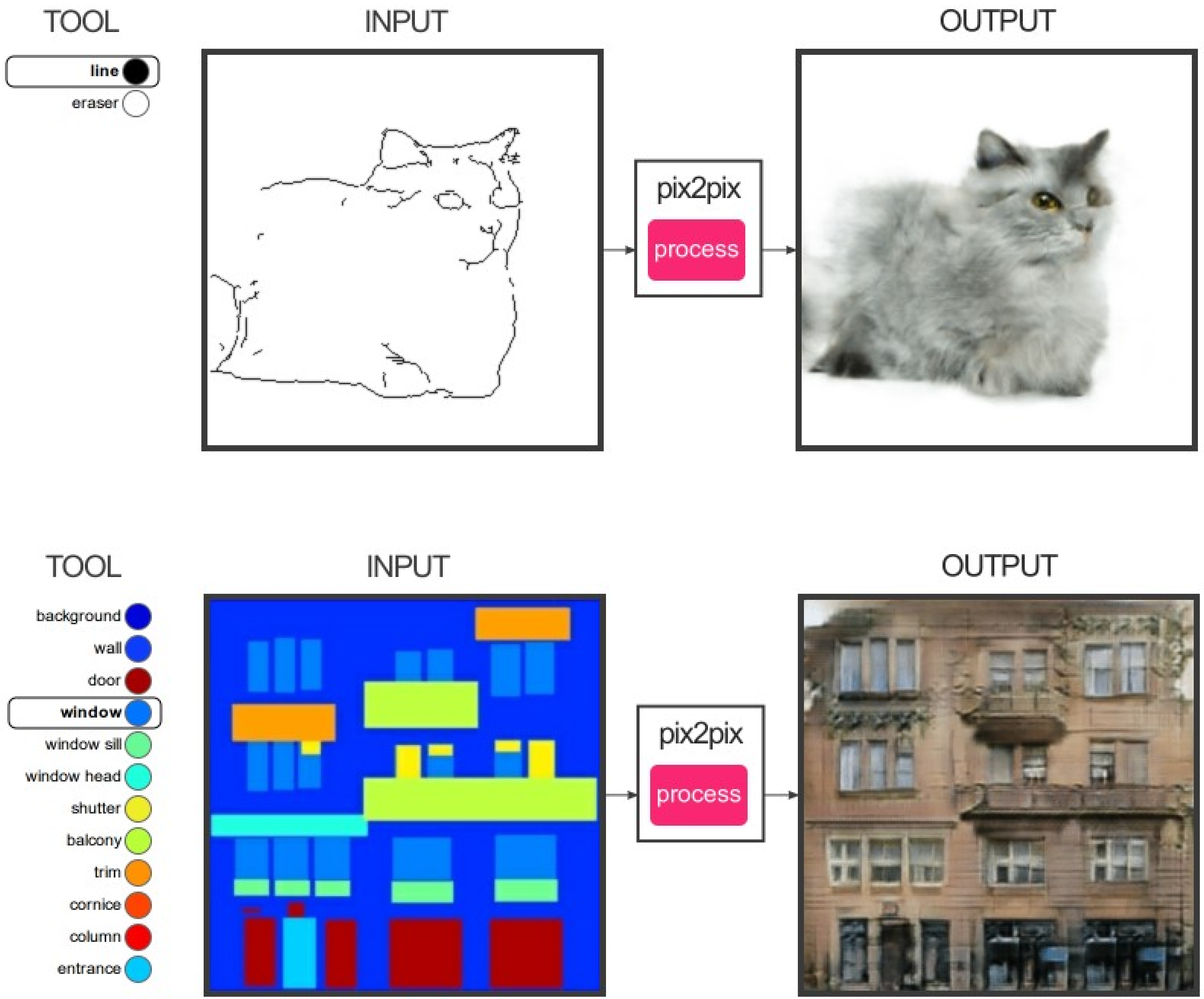
Real noisy photos

Input	Output
	
	


(a)
(b)

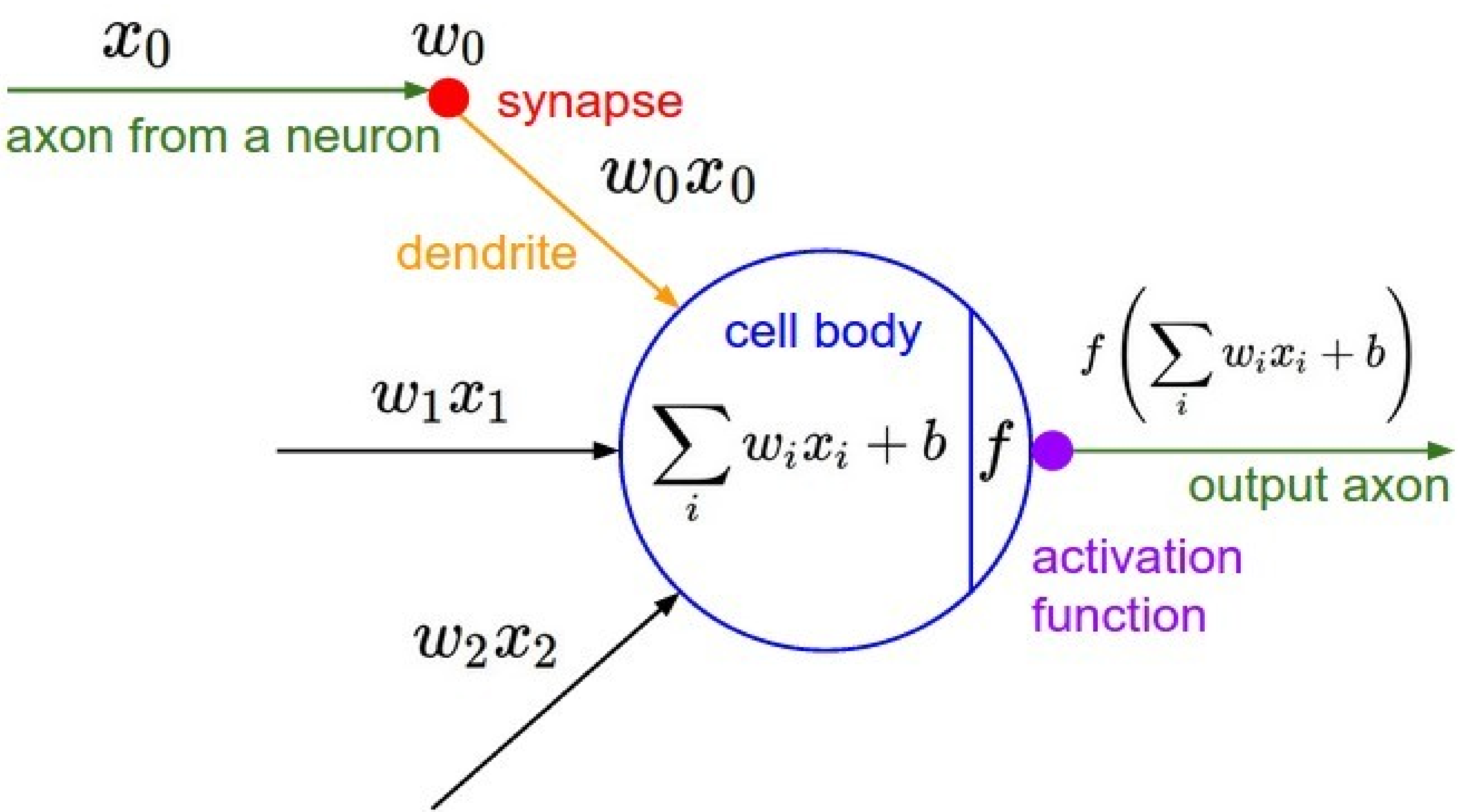


Caption	Generated Images
the flower shown has yellow anther red pistil and bright red petals	
this flower has petals that are yellow, white and purple and has dark lines	
the petals on this flower are white with a yellow center	
this flower has a lot of small round pink petals.	
this flower is orange in color, and has petals that are ruffled and rounded.	
the flower has yellow petals and the center of it is brown	


TOOL
line
eraser
INPUT
pix2pix
process
OUTPUT
TOOL
background
wall
door
window
window sill
window head
shutter
balcony
trim
cornice
column
entrance
INPUT
pix2pix
process
OUTPUT

What is a neural network?

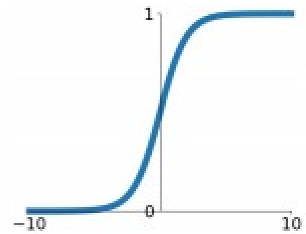
Single Neuron



Activation functions

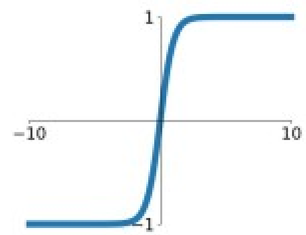
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



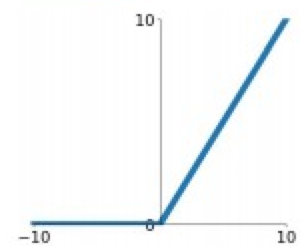
tanh

$\tanh(x)$



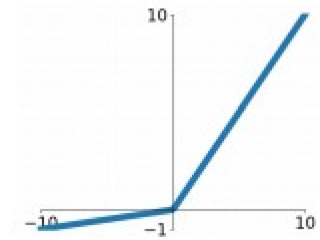
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

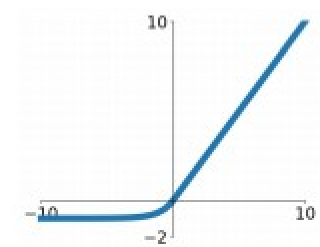


Maxout

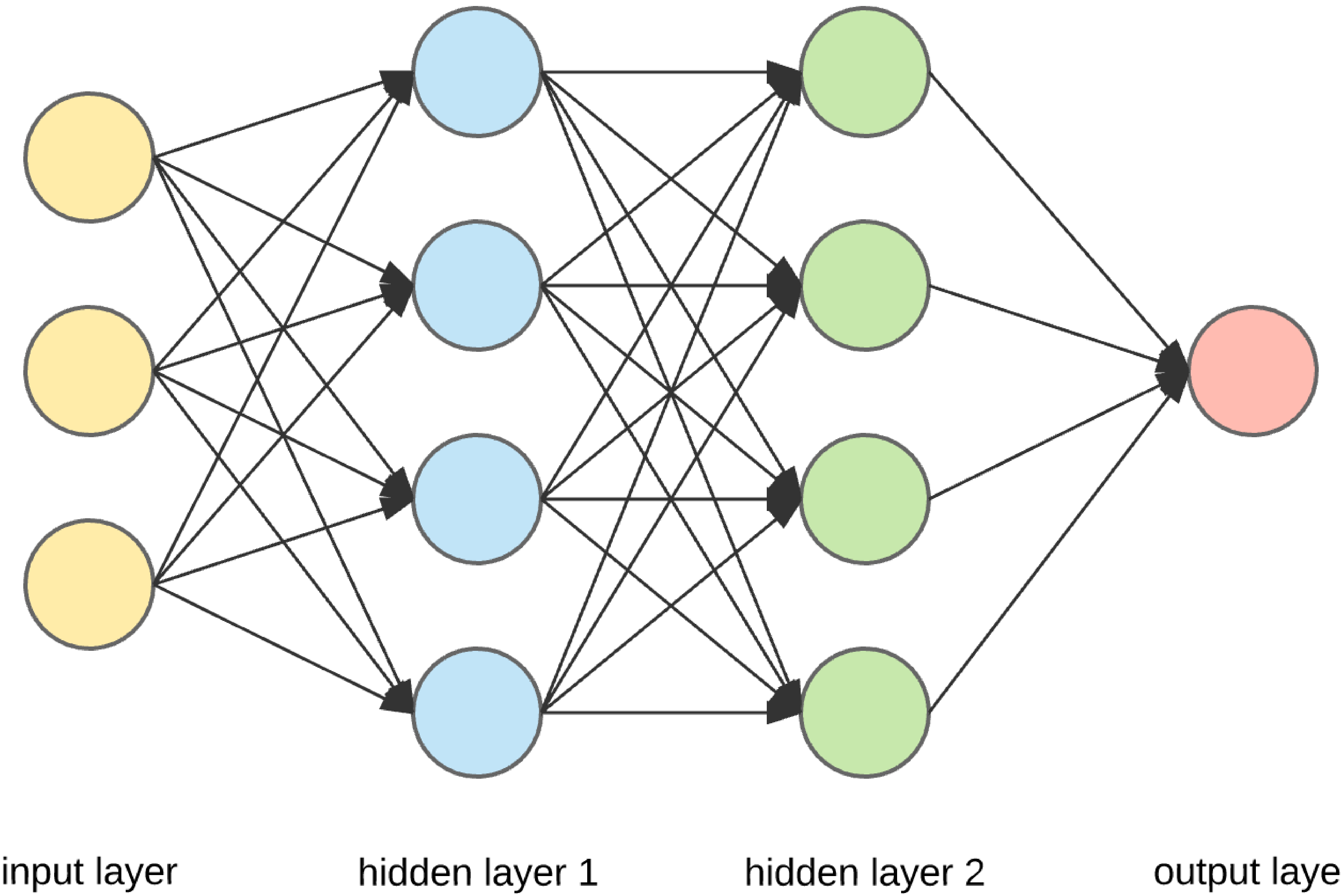
$\max(w_1^T x + b_1, w_2^T x + b_2)$

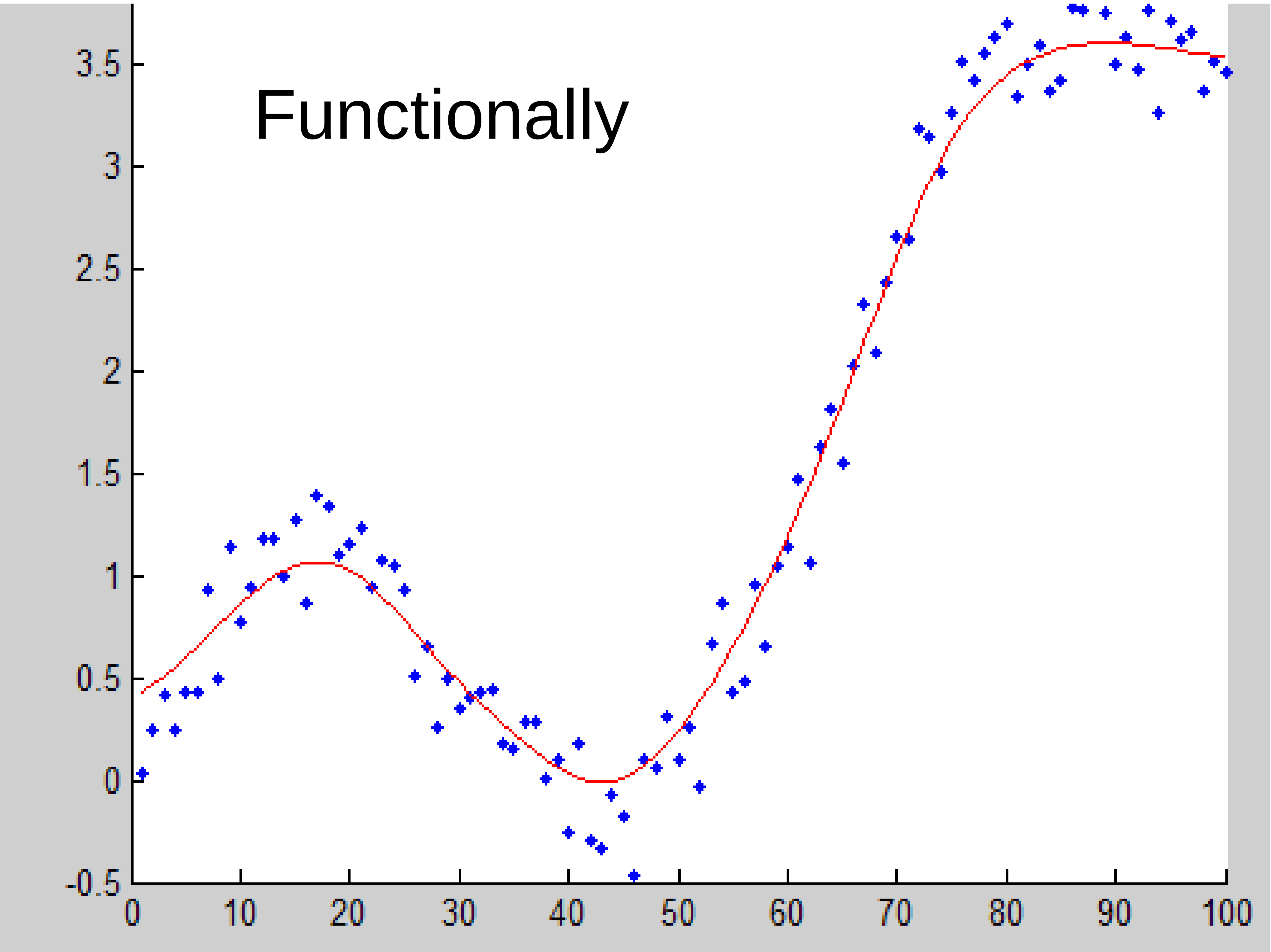
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Structurally




Functionally
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
0
10
20
30
40
50
60
70
80
90
100

Functionally, neural network is a gargantuan interpolator-approximator with millions internal parameters

Example: AlexNet, 62,378,344 parameters

Where do we get the points to approximate?

Three ways (maybe, you'll be the one to invent more)

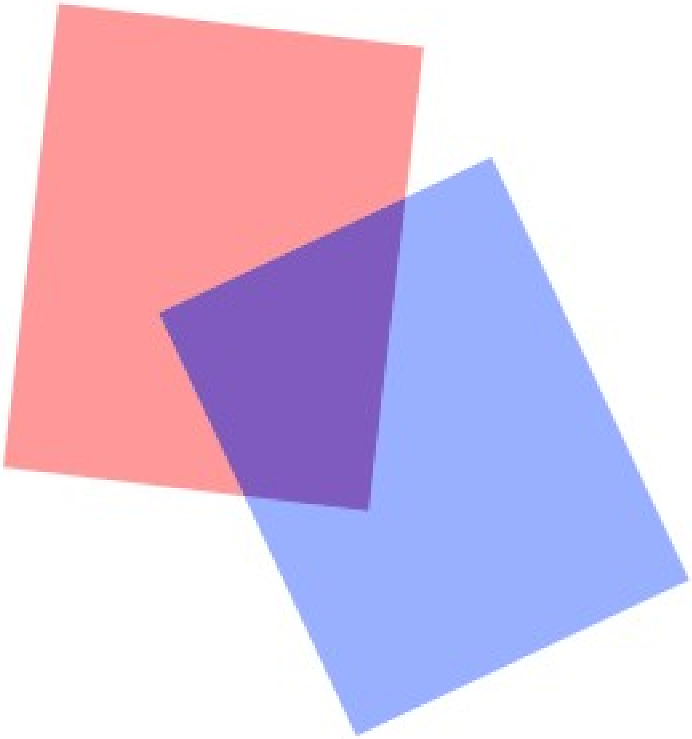
- We define the desired output ourselves – Supervised learning (classification, etc.)
- We cannot define the desired output, but we can tell how bad is the given output – Unsupervised learning (clustering, etc.)
- We allow NN to interact with the environment and assess the consequences – Reinforcement learning (play Atari game, etc.)

How do we know, neural network does its job good?

How do we know, it does what we want?

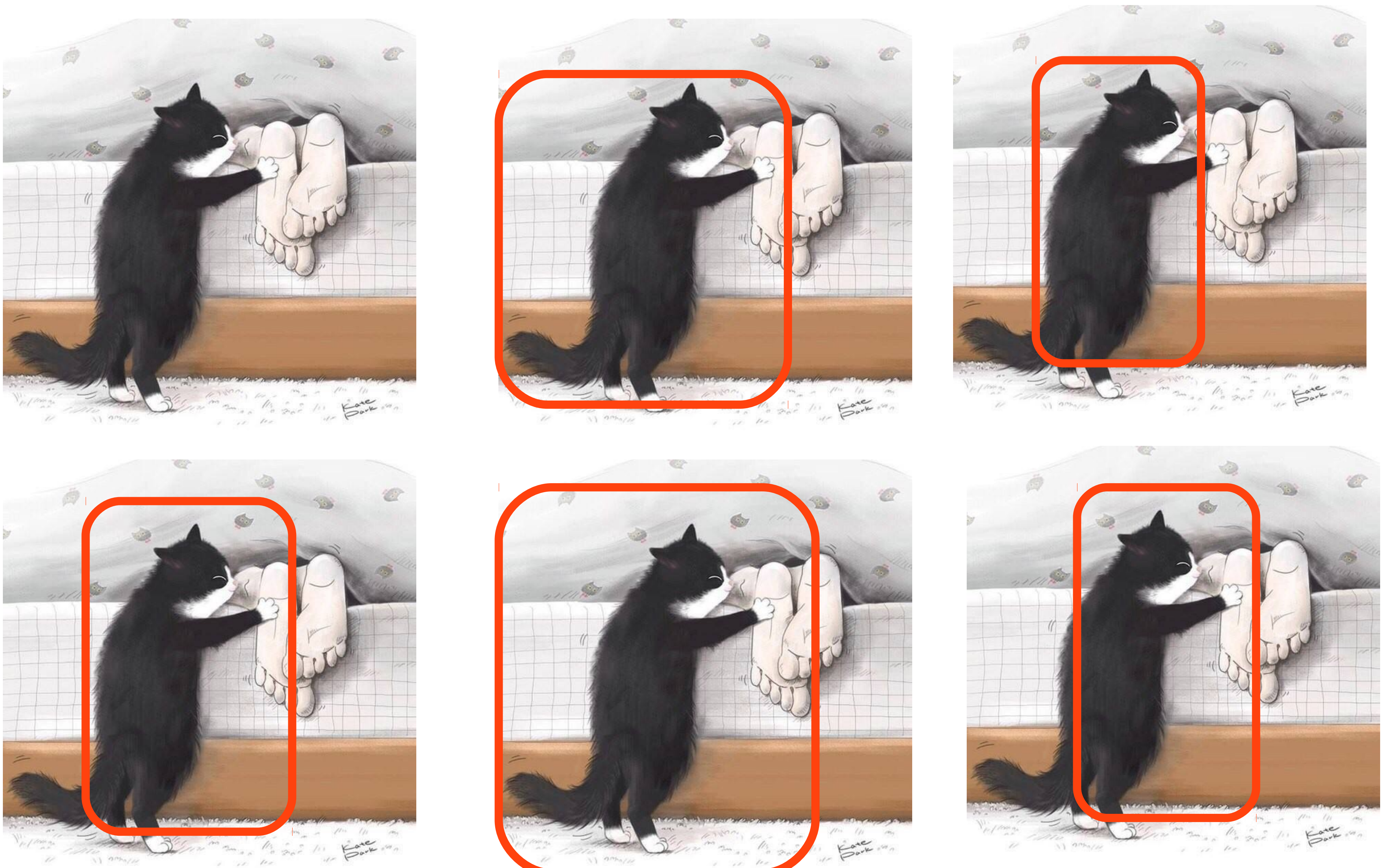
Cost Function
also known as
Loss Function

to the rescue!

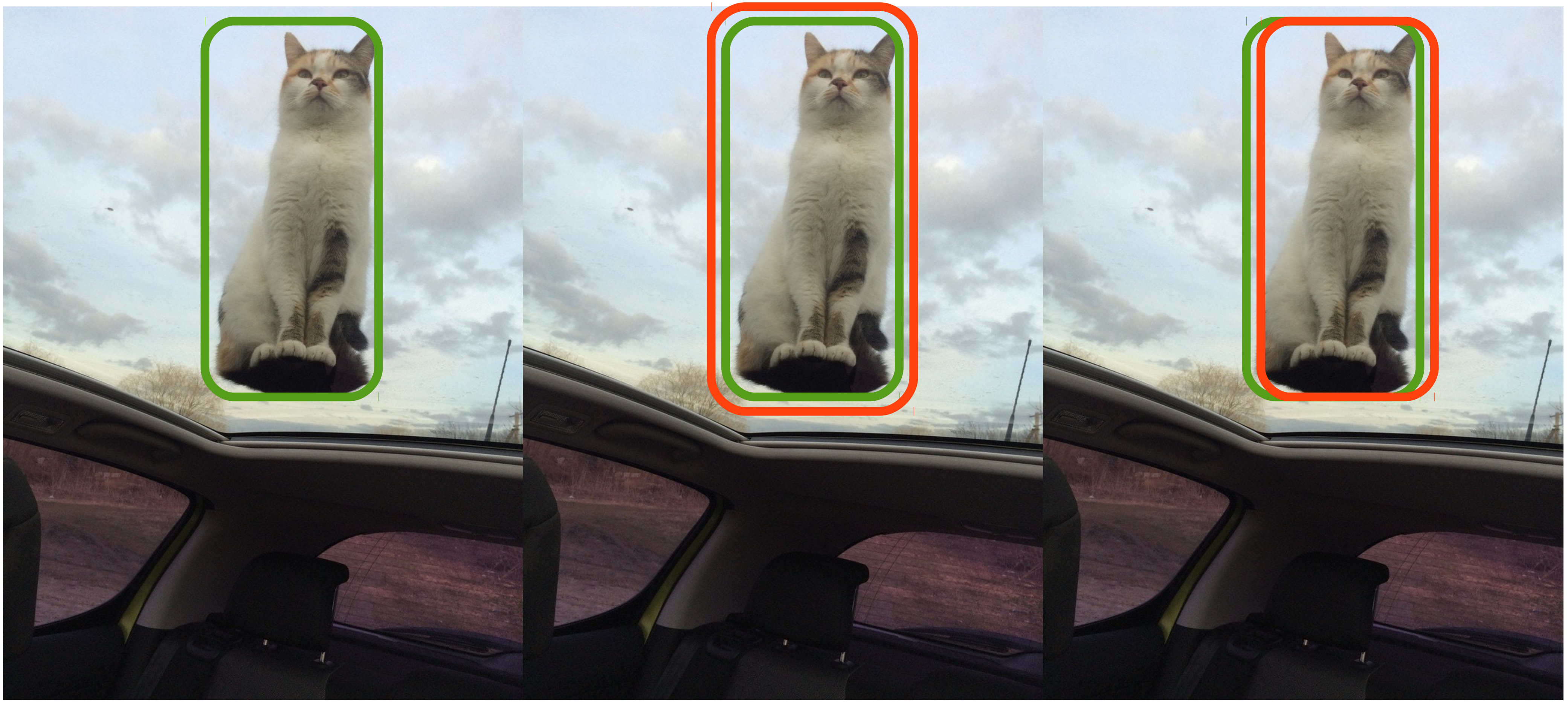


How to formalize
that two rectangles
match?

What is supposed to be “good”?



Now you have ground-truth



Which detect is better?

Where do we get internal parameters?

We train neural network

How do we train neural network?

Methodological point of view

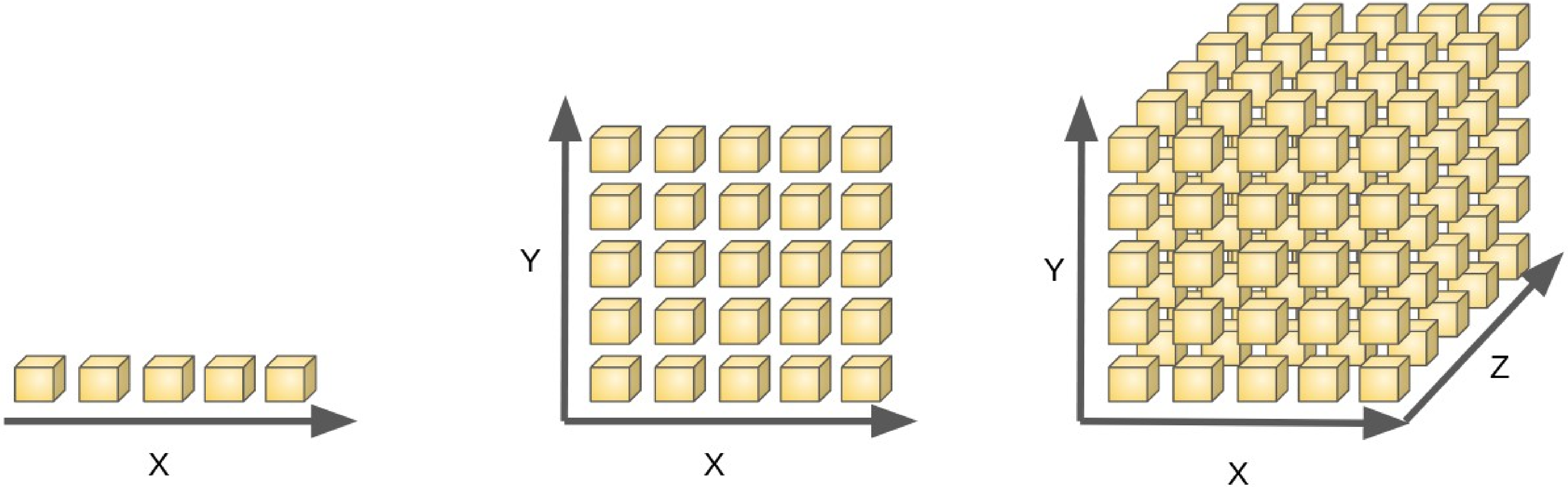
- Supervised learning (classification, etc.)
- Unsupervised learning (clustering, etc.)
- Reinforcement learning (play Atari game, etc.)

Implementation point of view

Loss function, the formalization of how good NN performs, should be minimized (we could define “gain” function and maximize it)

What minimization methods do we know? What is the suitable one?

Parameters... Parameters everywhere...



Curse of dimensions

Function of millions of arguments...

How do we minimize it?

What properties can we rely on?
(global or local)

We can rely on local properties only due to curse of dimensions

So neural network is just a function

$$f_{\alpha_1,\alpha_2,\ldots,\alpha_n}(x_1, x_2, \ldots, x_m)$$

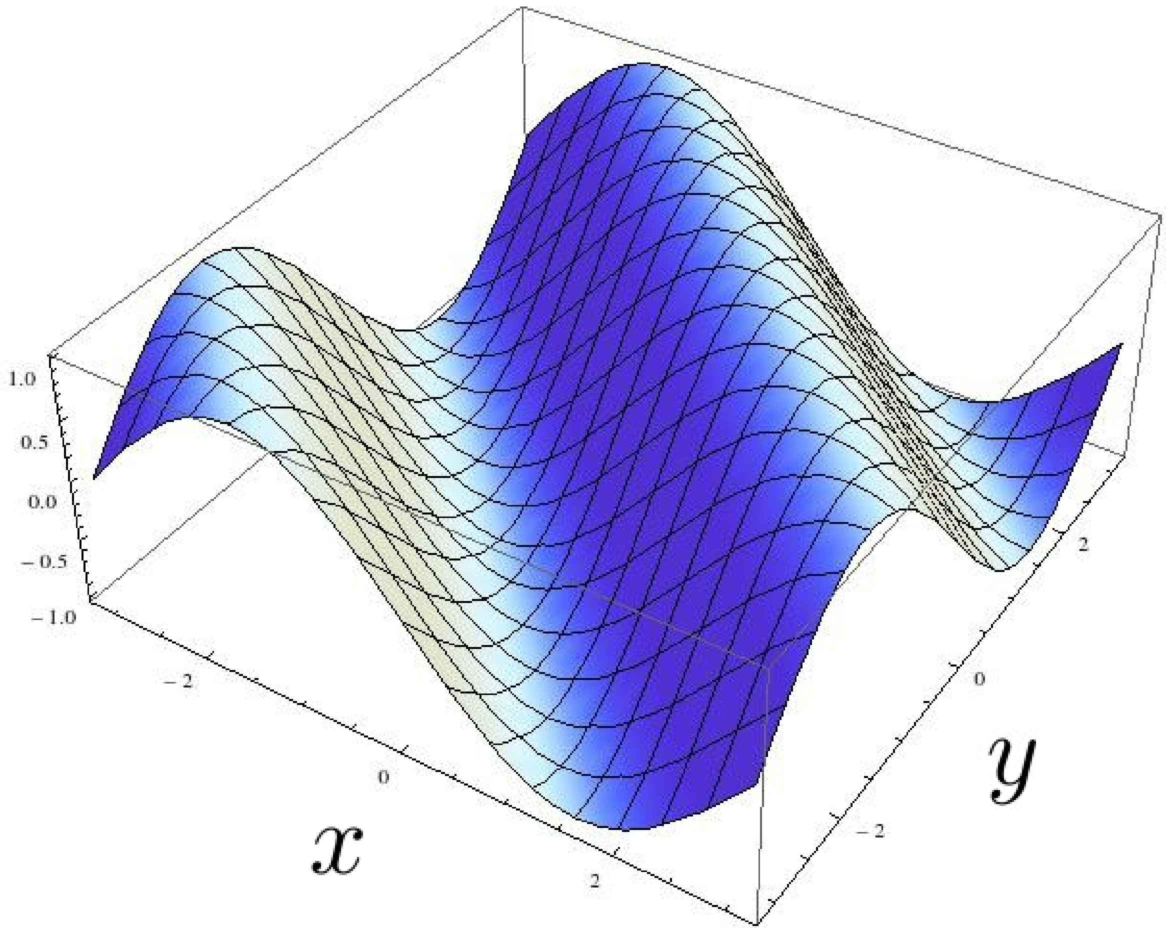
We can rely on local properties only due to curse of dimensions

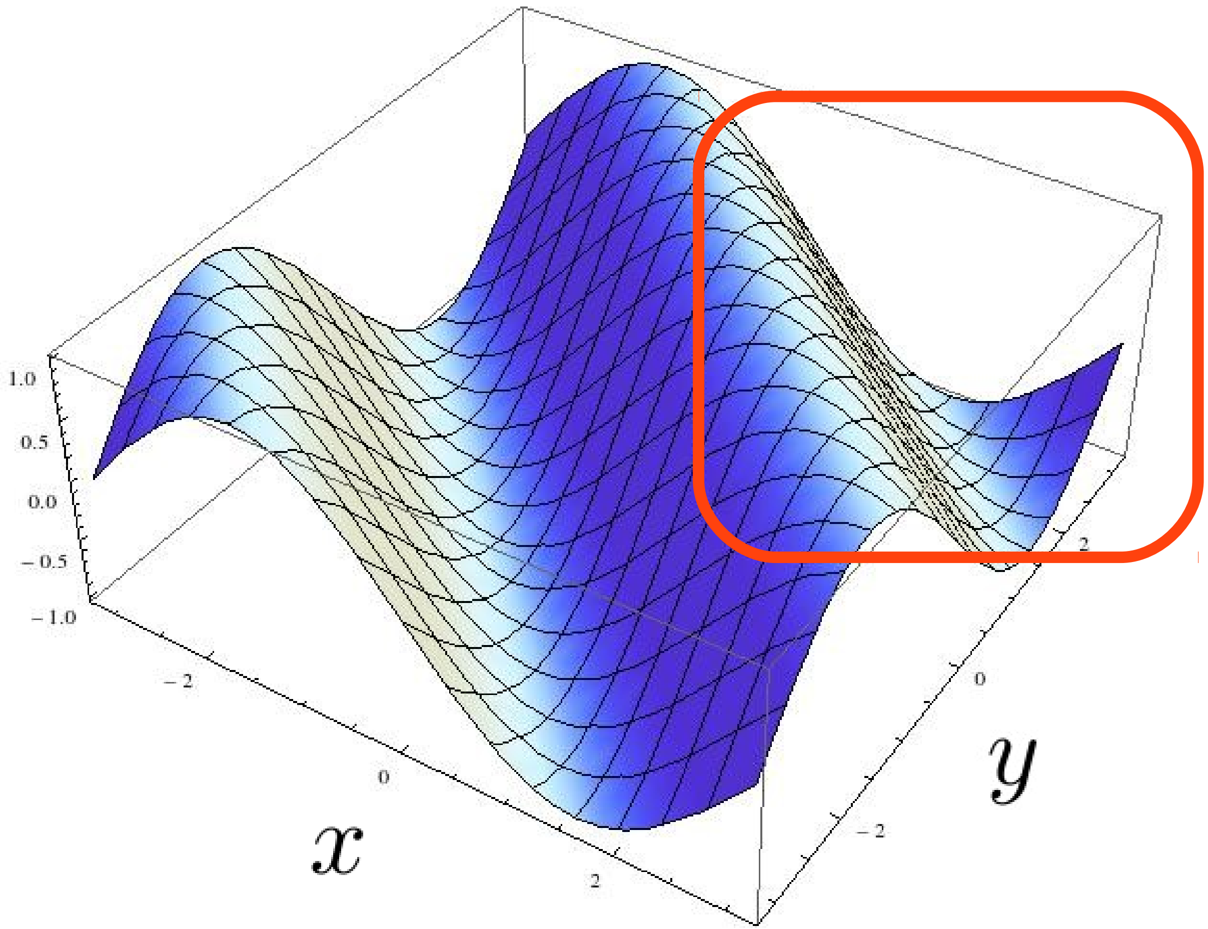
So neural network is just a function

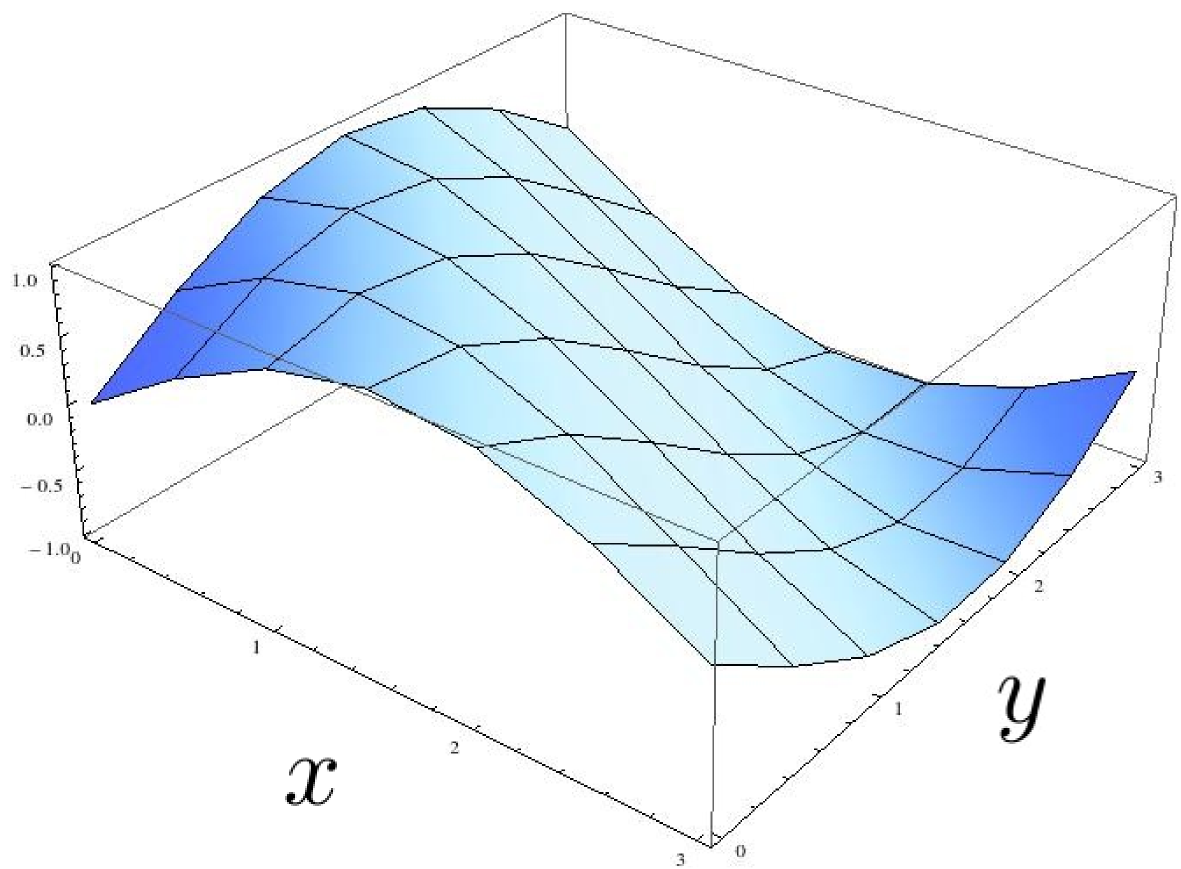
$$f_{\alpha_1,\alpha_2,\ldots,\alpha_n}(x_1, x_2, \ldots, x_m)$$

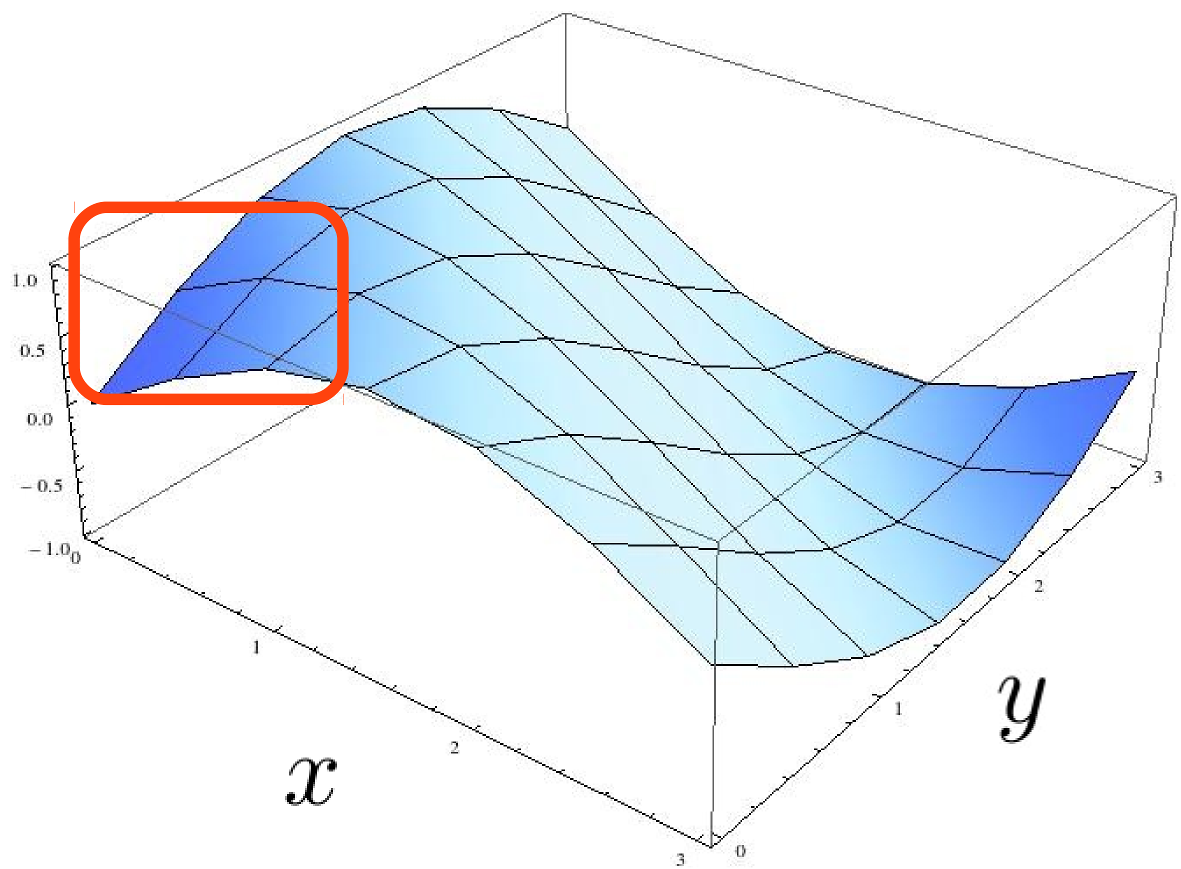
It is philosophical question what to consider parameters and what should be arguments

$$f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m)$$


1.0
0.5
0.0
−0.5
−1.0
−2
0
2
x
−2
0
2
y

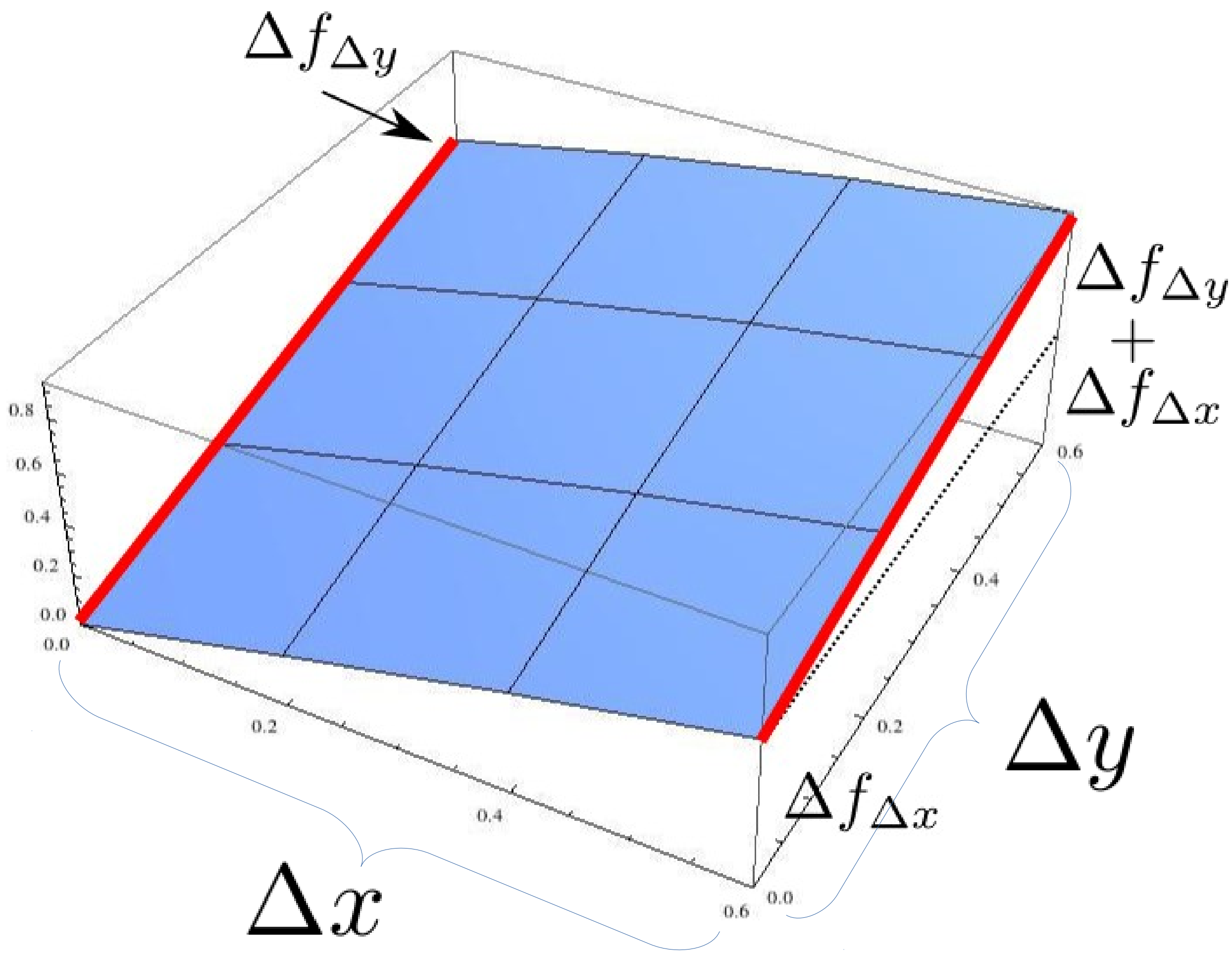

1.0
0.5
0.0
− 0.5
− 1.0
− 2
0
2
x
− 2
0
2
y


1.0
0.5
0.0
− 0.5
− 1.0
0
1
2
3
x
0
1
2
3
y



1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
x
0
1
2
3
y

Δx and Δy not equal in general!



Red lines we assume to be parallel

Now consider small deviation

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \frac{\partial f}{\partial \alpha_1} \Delta \alpha_1 + \frac{\partial f}{\partial \alpha_2} \Delta \alpha_2 + \cdots + \frac{\partial f}{\partial \alpha_n} \Delta \alpha_n$$

Now consider small deviation

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \frac{\partial f}{\partial \alpha_1}\Delta\alpha_1 + \frac{\partial f}{\partial \alpha_2}\Delta\alpha_2 + \cdots + \frac{\partial f}{\partial \alpha_n}\Delta\alpha_n$$

Do a trick

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \overrightarrow{\left\{\frac{\partial f}{\partial \alpha_1}; \frac{\partial f}{\partial \alpha_2}; \ldots; \frac{\partial f}{\partial \alpha_n}\right\}} \cdot \overrightarrow{\{\Delta\alpha_1; \Delta\alpha_2; \ldots; \Delta\alpha_n\}} = \vec{\nabla}_\alpha f \cdot \vec{\Delta}_\alpha$$

Now consider small deviation

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \frac{\partial f}{\partial \alpha_1}\Delta\alpha_1 + \frac{\partial f}{\partial \alpha_2}\Delta\alpha_2 + \cdots + \frac{\partial f}{\partial \alpha_n}\Delta\alpha_n$$

Do a trick

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \overrightarrow{\left\{\frac{\partial f}{\partial \alpha_1}; \frac{\partial f}{\partial \alpha_2}; \ldots; \frac{\partial f}{\partial \alpha_n}\right\}} \cdot \overrightarrow{\{\Delta\alpha_1; \Delta\alpha_2; \ldots; \Delta\alpha_n\}} = \vec{\nabla}_\alpha f \cdot \vec{\Delta}_\alpha$$

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = |\vec{\nabla}_\alpha f| \; \underbrace{|\vec{\Delta}_\alpha|}_{\lambda} \cos(\varphi)$$

Now consider small deviation

$$\Delta f(\alpha_1,\alpha_2,\ldots,\alpha_n;x_1,x_2,\ldots,x_m) = \frac{\partial f}{\partial \alpha_1}\Delta\alpha_1 + \frac{\partial f}{\partial \alpha_2}\Delta\alpha_2 + \cdots + \frac{\partial f}{\partial \alpha_n}\Delta\alpha_n$$

Do a trick

$$\Delta f(\alpha_1,\alpha_2,\ldots,\alpha_n;x_1,x_2,\ldots,x_m) = \overrightarrow{\left\{\frac{\partial f}{\partial \alpha_1};\frac{\partial f}{\partial \alpha_2};\ldots;\frac{\partial f}{\partial \alpha_n}\right\}} \cdot \overrightarrow{\{\Delta\alpha_1;\Delta\alpha_2;\ldots;\Delta\alpha_n\}} = \vec{\nabla}_\alpha f \cdot \vec{\Delta}_\alpha$$

$$\Delta f(\alpha_1,\alpha_2,\ldots,\alpha_n;x_1,x_2,\ldots,x_m) = |\vec{\nabla}_\alpha f| \underbrace{|\vec{\Delta}_\alpha|}_{\lambda} \cos(\varphi)$$

How to make our function minimal?

Now consider small deviation

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \frac{\partial f}{\partial \alpha_1}\Delta\alpha_1 + \frac{\partial f}{\partial \alpha_2}\Delta\alpha_2 + \cdots + \frac{\partial f}{\partial \alpha_n}\Delta\alpha_n$$

Do a trick

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \overrightarrow{\left\{\frac{\partial f}{\partial \alpha_1}; \frac{\partial f}{\partial \alpha_2}; \ldots; \frac{\partial f}{\partial \alpha_n}\right\}} \cdot \overrightarrow{\{\Delta\alpha_1; \Delta\alpha_2; \ldots; \Delta\alpha_n\}} = \vec{\nabla}_\alpha f \cdot \vec{\Delta}_\alpha$$

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = |\vec{\nabla}_\alpha f| \; \underbrace{|\vec{\Delta}_\alpha|}_{\lambda} \cos(\varphi)$$

How to make our function minimal?

$$\cos(\varphi) = -1; \qquad \lambda > 0$$

Now consider small deviation

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \frac{\partial f}{\partial \alpha_1}\Delta\alpha_1 + \frac{\partial f}{\partial \alpha_2}\Delta\alpha_2 + \cdots + \frac{\partial f}{\partial \alpha_n}\Delta\alpha_n$$

Do a trick

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = \overrightarrow{\left\{\frac{\partial f}{\partial \alpha_1}; \frac{\partial f}{\partial \alpha_2}; \ldots; \frac{\partial f}{\partial \alpha_n}\right\}} \cdot \overrightarrow{\{\Delta\alpha_1; \Delta\alpha_2; \ldots; \Delta\alpha_n\}} = \vec{\nabla}_\alpha f \cdot \vec{\Delta}_\alpha$$

$$\Delta f(\alpha_1, \alpha_2, \ldots, \alpha_n; x_1, x_2, \ldots, x_m) = |\vec{\nabla}_\alpha f| \underbrace{|\vec{\Delta}_\alpha|}_{\lambda} \cos(\varphi)$$

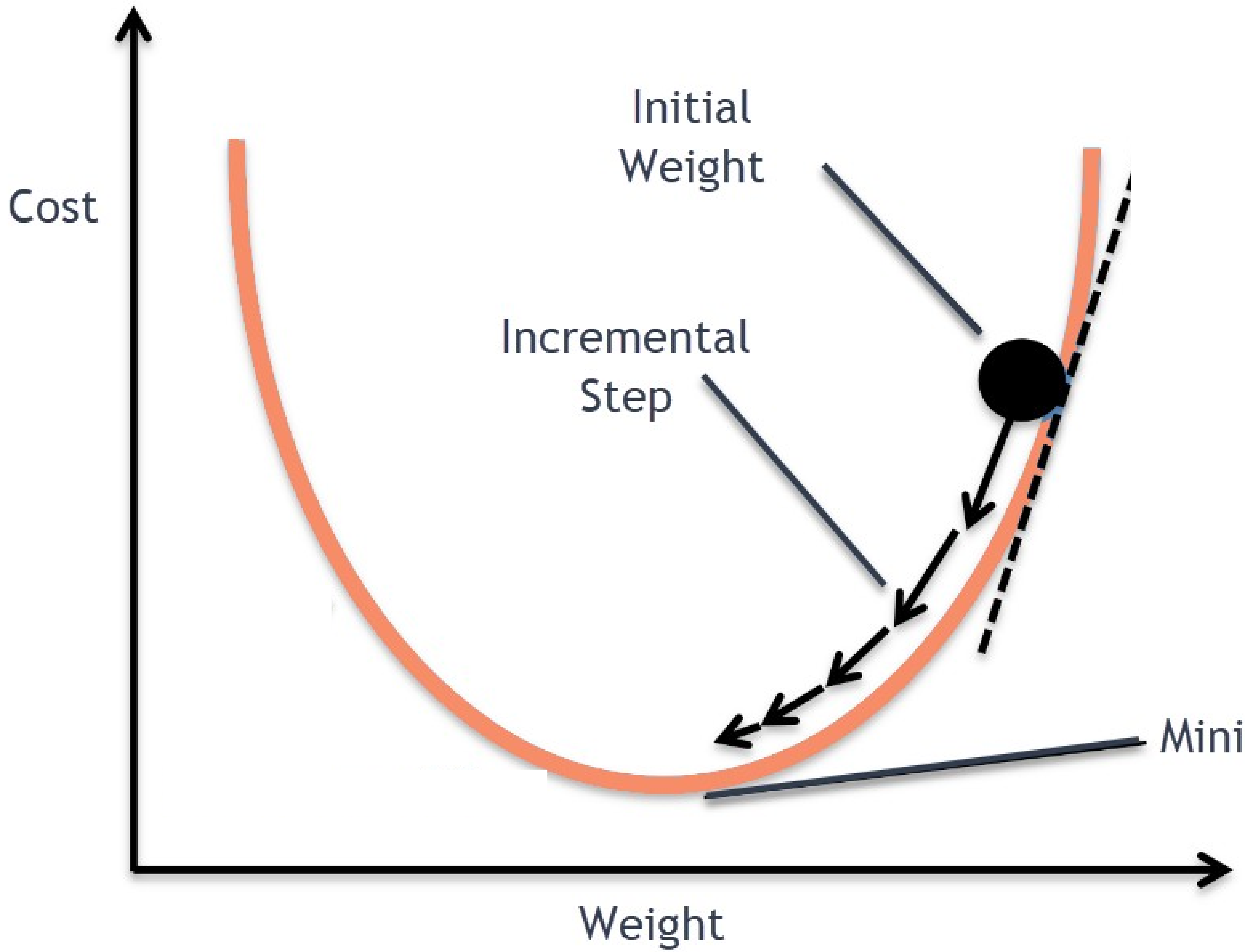
How to make our function minimal?

$$\cos(\varphi) = -1; \qquad \lambda > 0$$

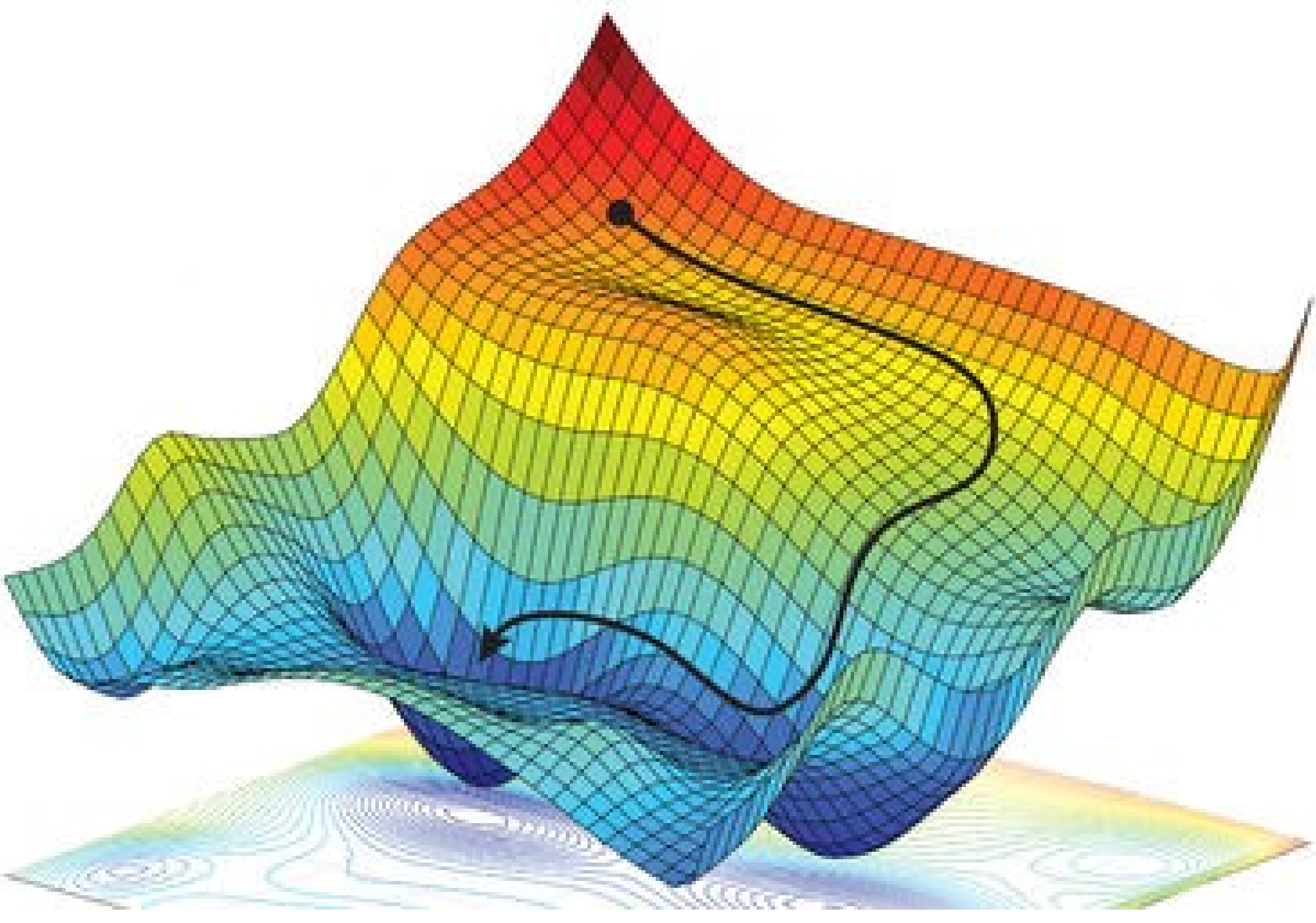
Here comes the idea of gradient descent

$$\vec{x}^{\,(n+1)} = \vec{x}^{\,(n)} - \lambda\, \vec{\nabla}_\alpha f\, \left(\vec{x}^{\,(n)}\right)$$

Gradient descent



Gradient descent



Modifications of gradient descent

- Momentum optimization
- Nesterov Momentum optimization
- AdaGrad
- RMSProp
- Adam optimization
- Learning rate scheduling

But I can't write NN as a single function
(well... I can but it will take forever)

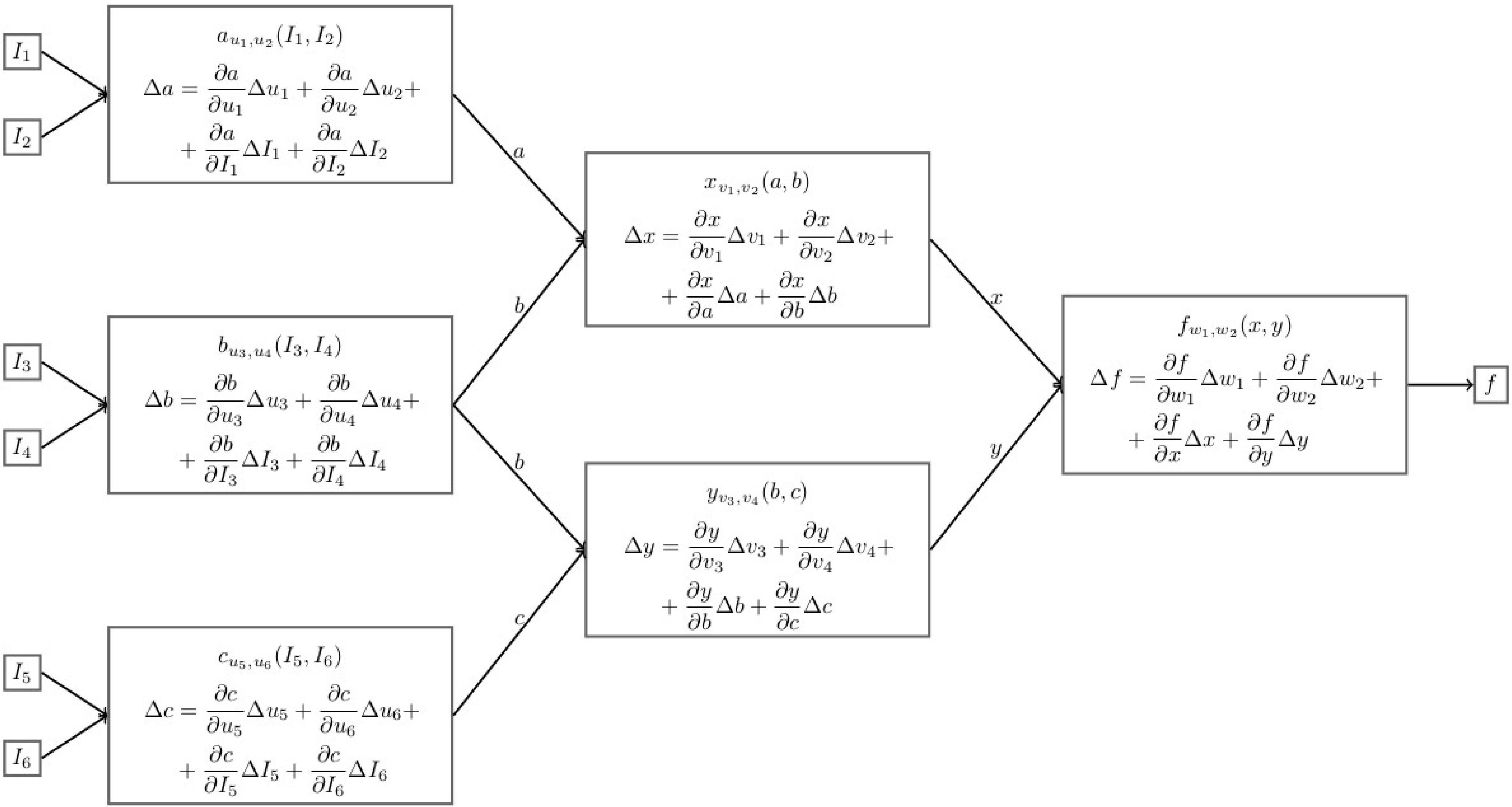
How do I find gradient?

Backpropagation

Backpropagation is based on the chain rule

Note: all derivatives are meant to be calculated at a certain point – they are numbers!

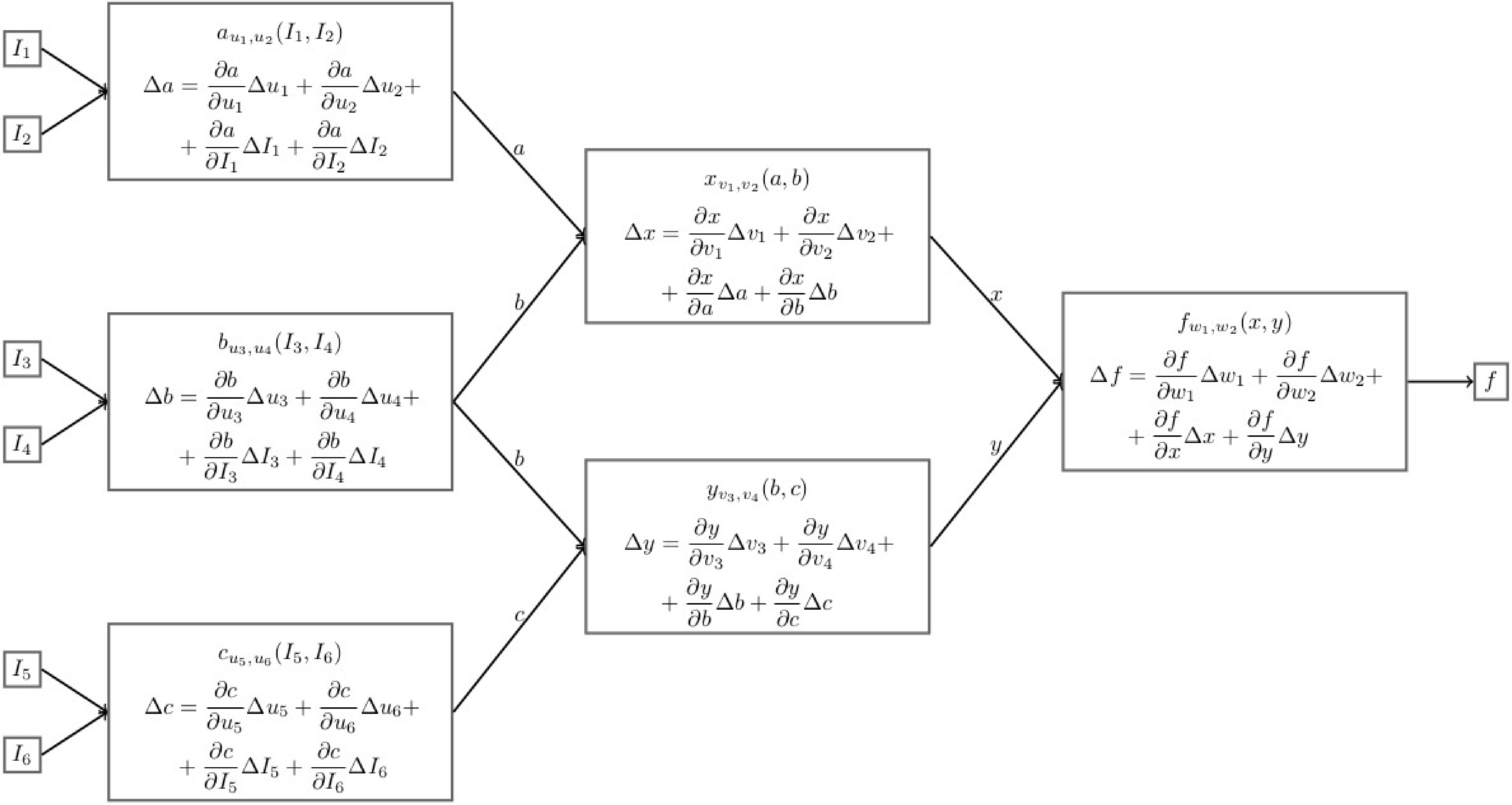
$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$



Note: all derivatives are meant to be calculated at a certain point – they are numbers!

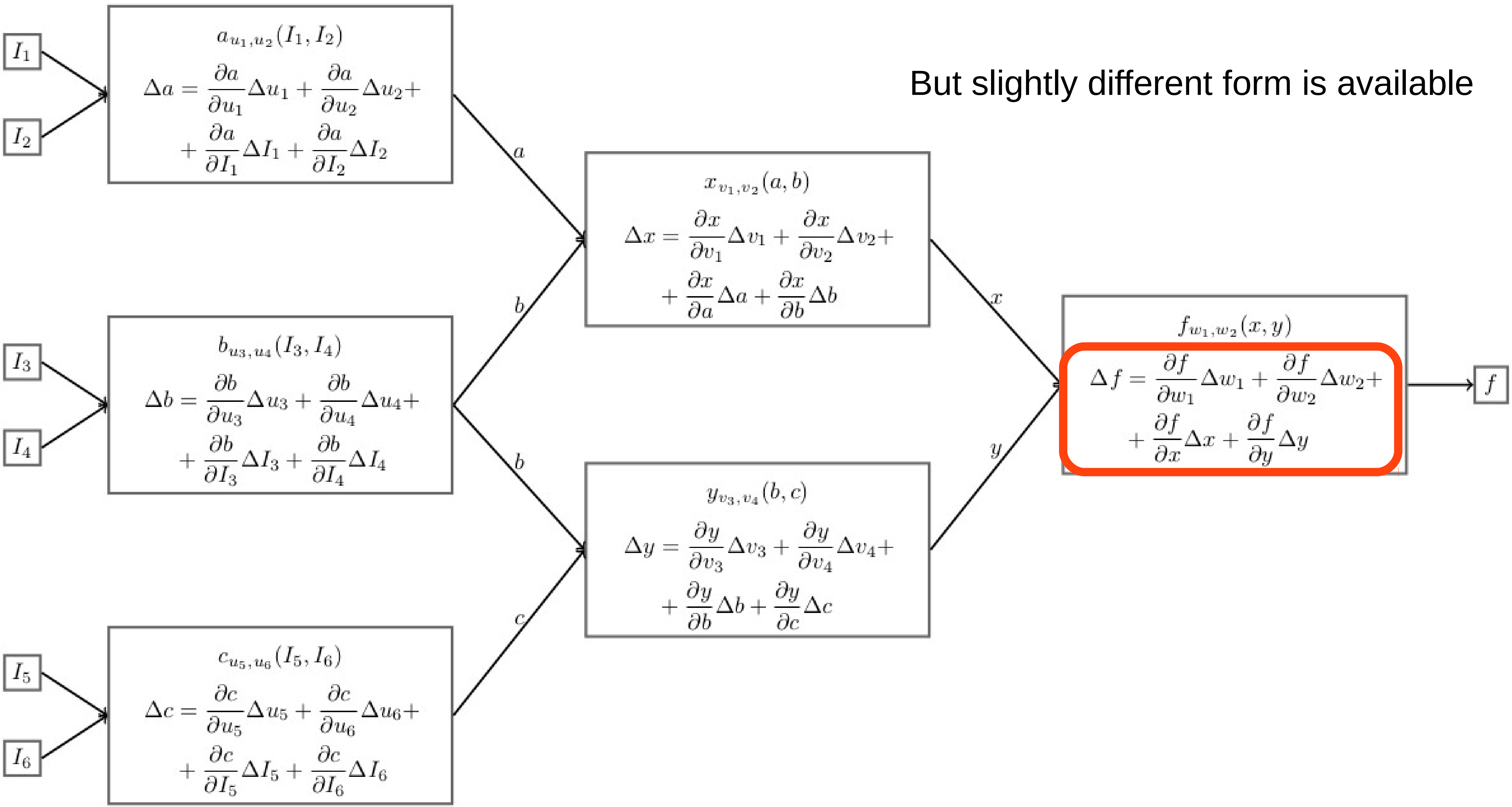
$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$

I want change of the loss function dependent on all network's parameters



Note: all derivatives are meant to be calculated at a certain point – they are numbers!

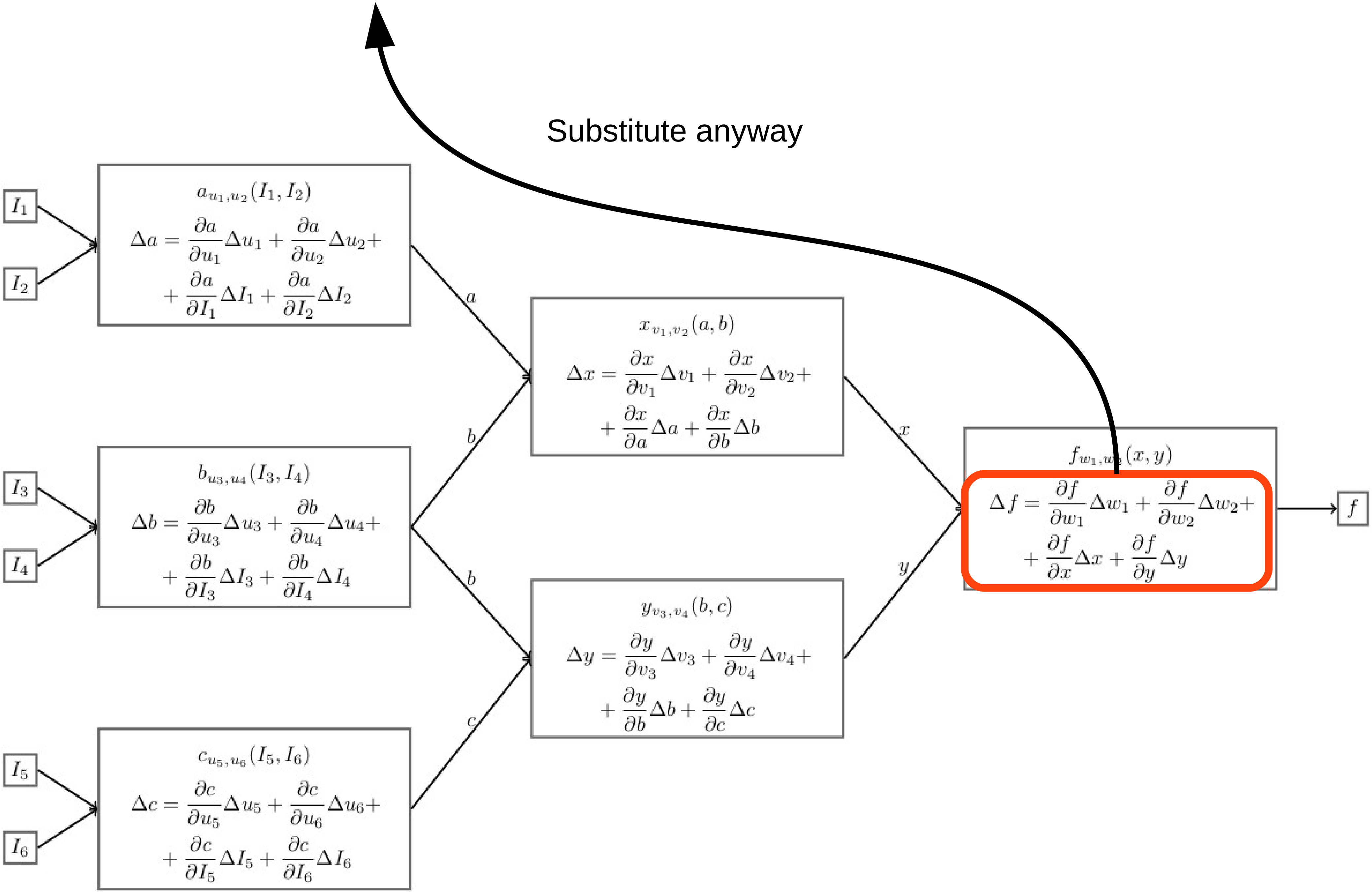
$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$



But slightly different form is available

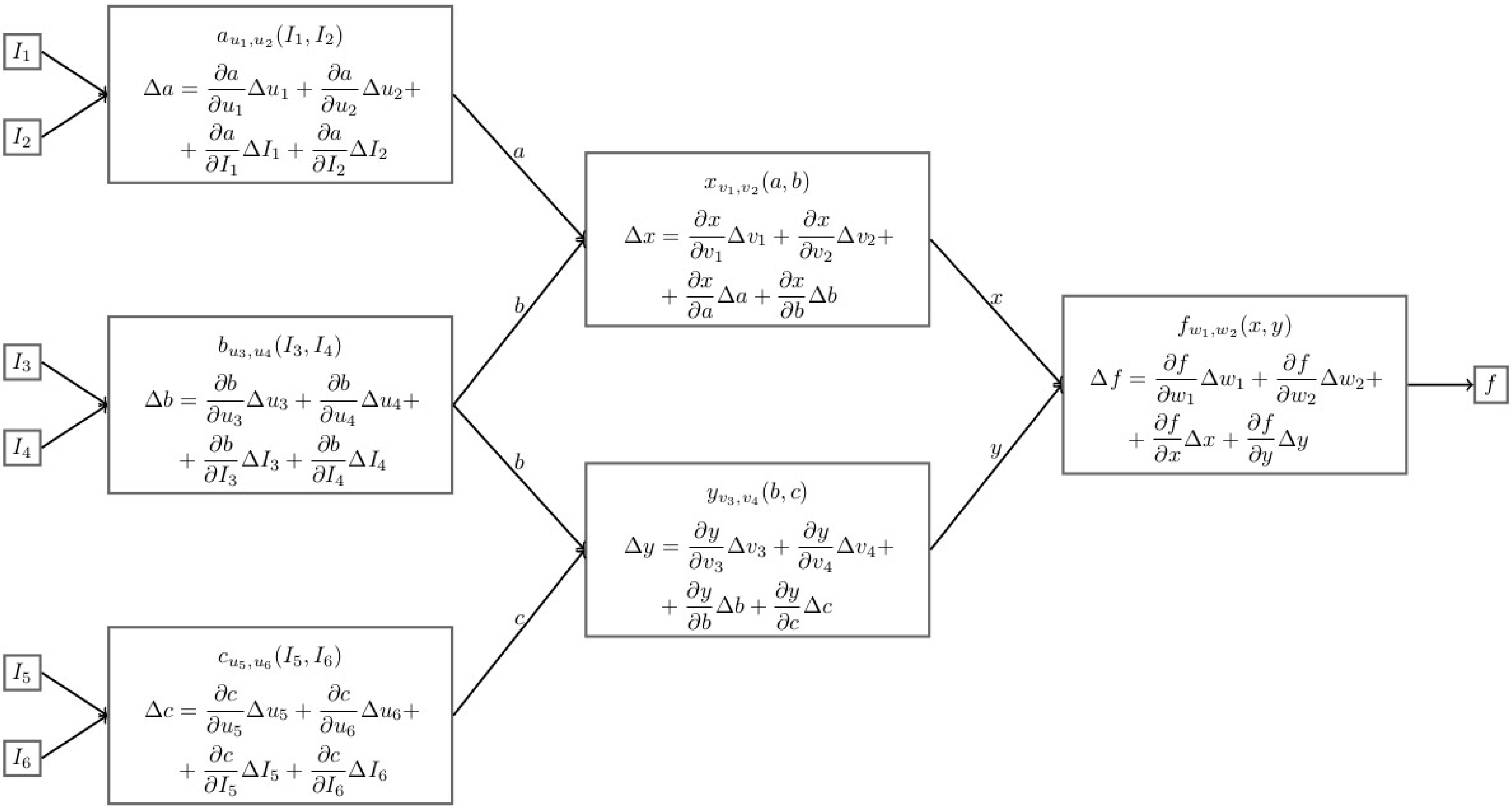
Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$



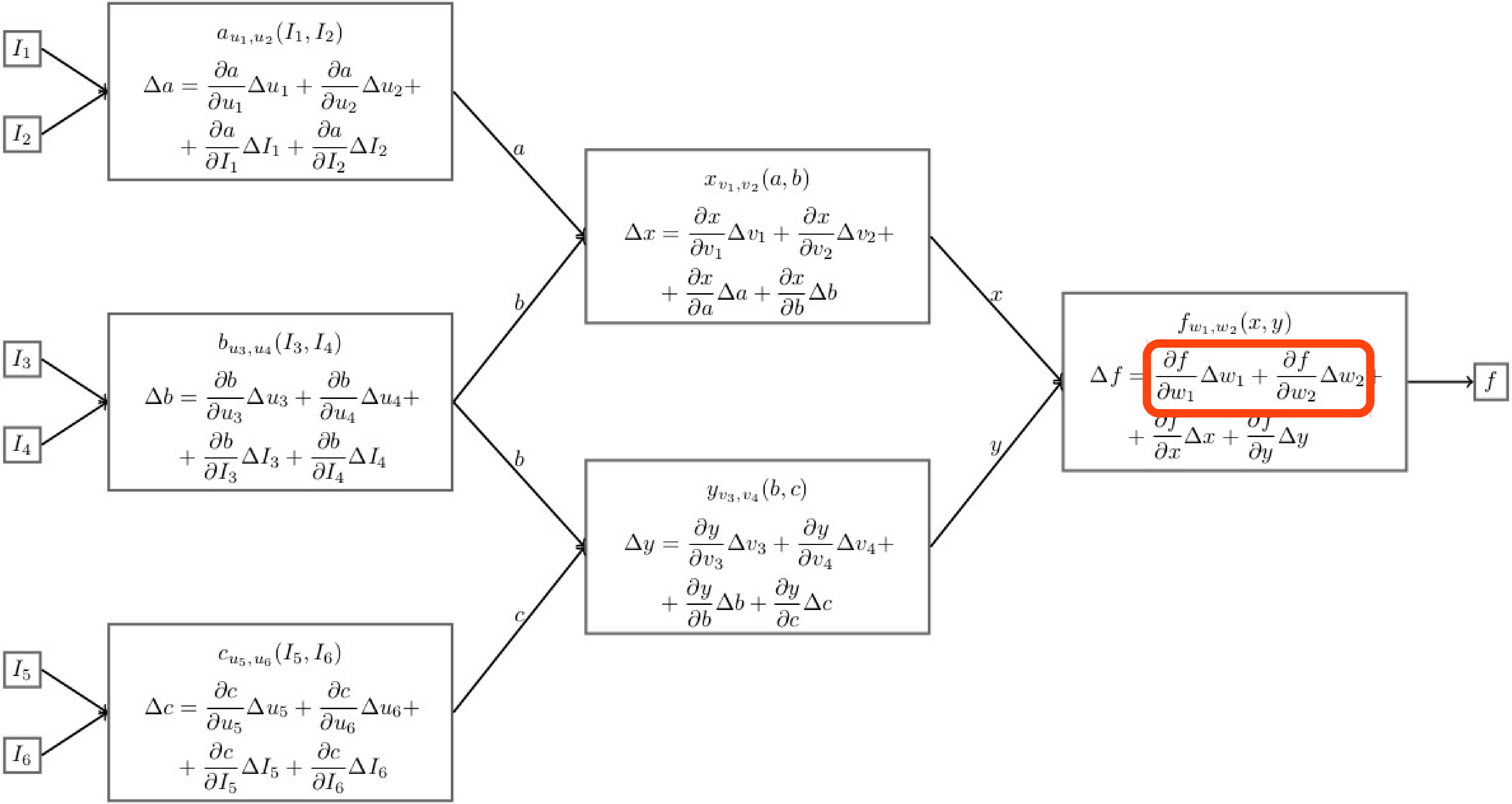
Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 +$$
$$+ \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$



Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 +$$
$$+ \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$

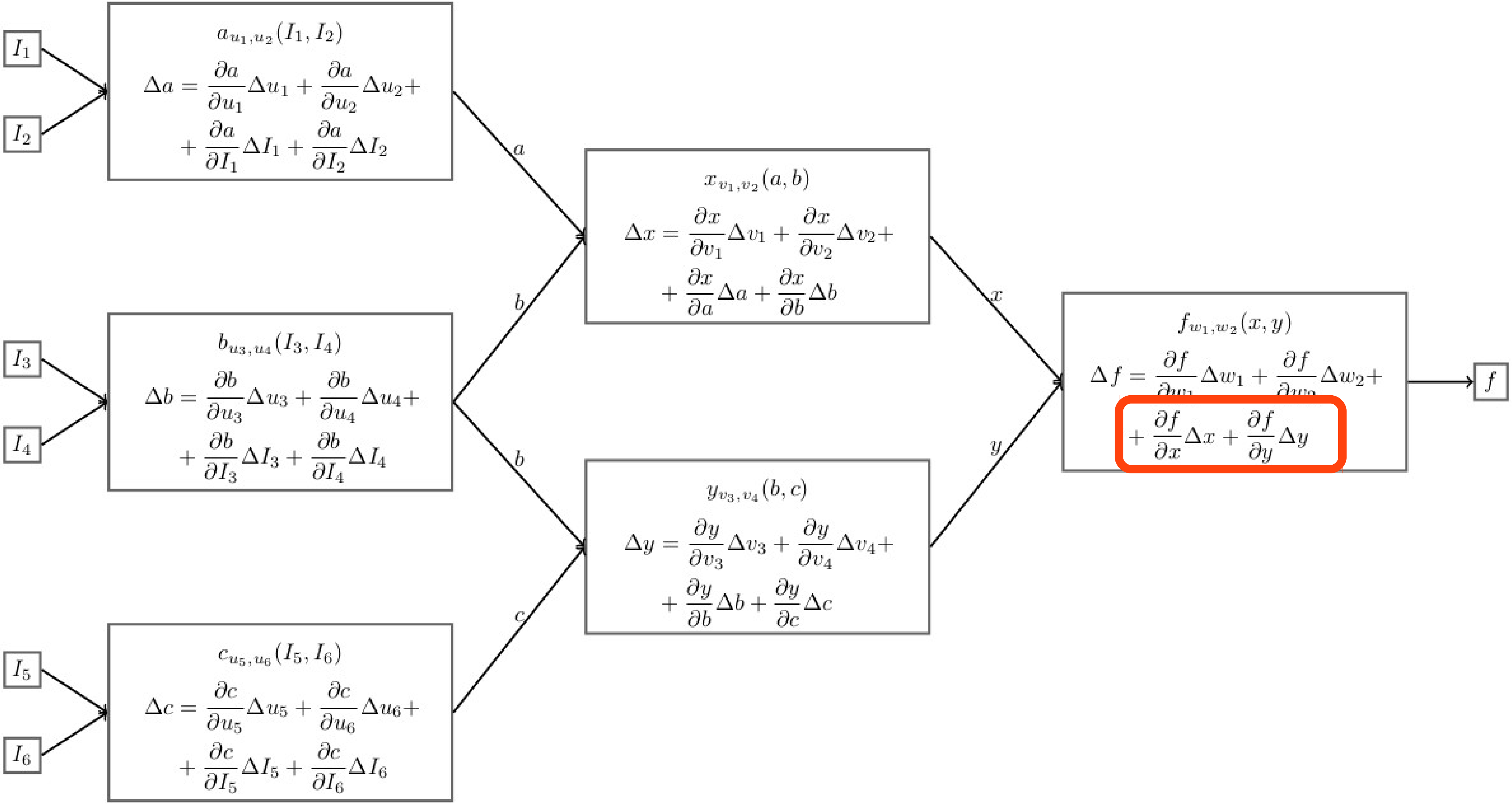


Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$

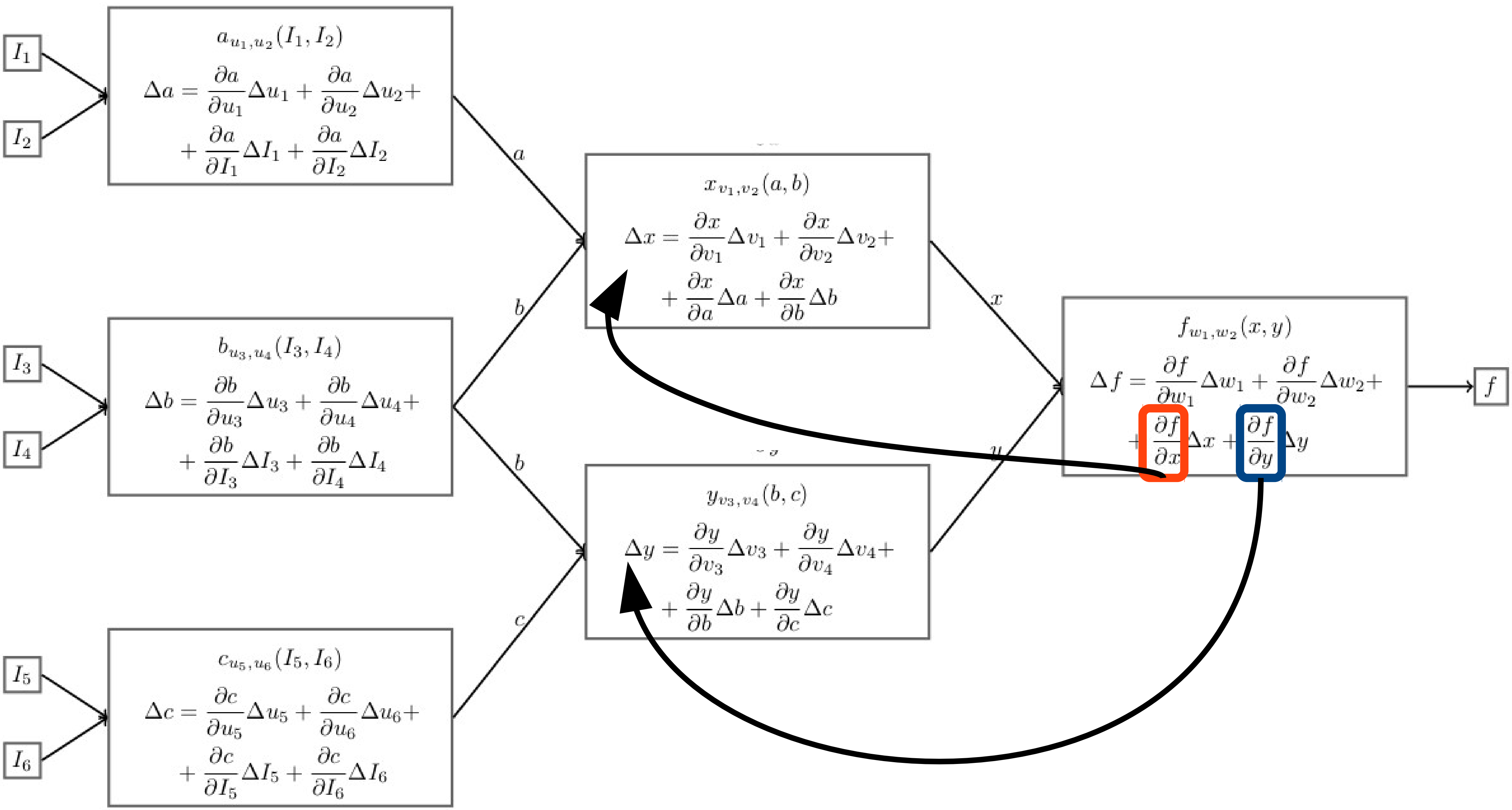
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 +$$

$$+ \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$



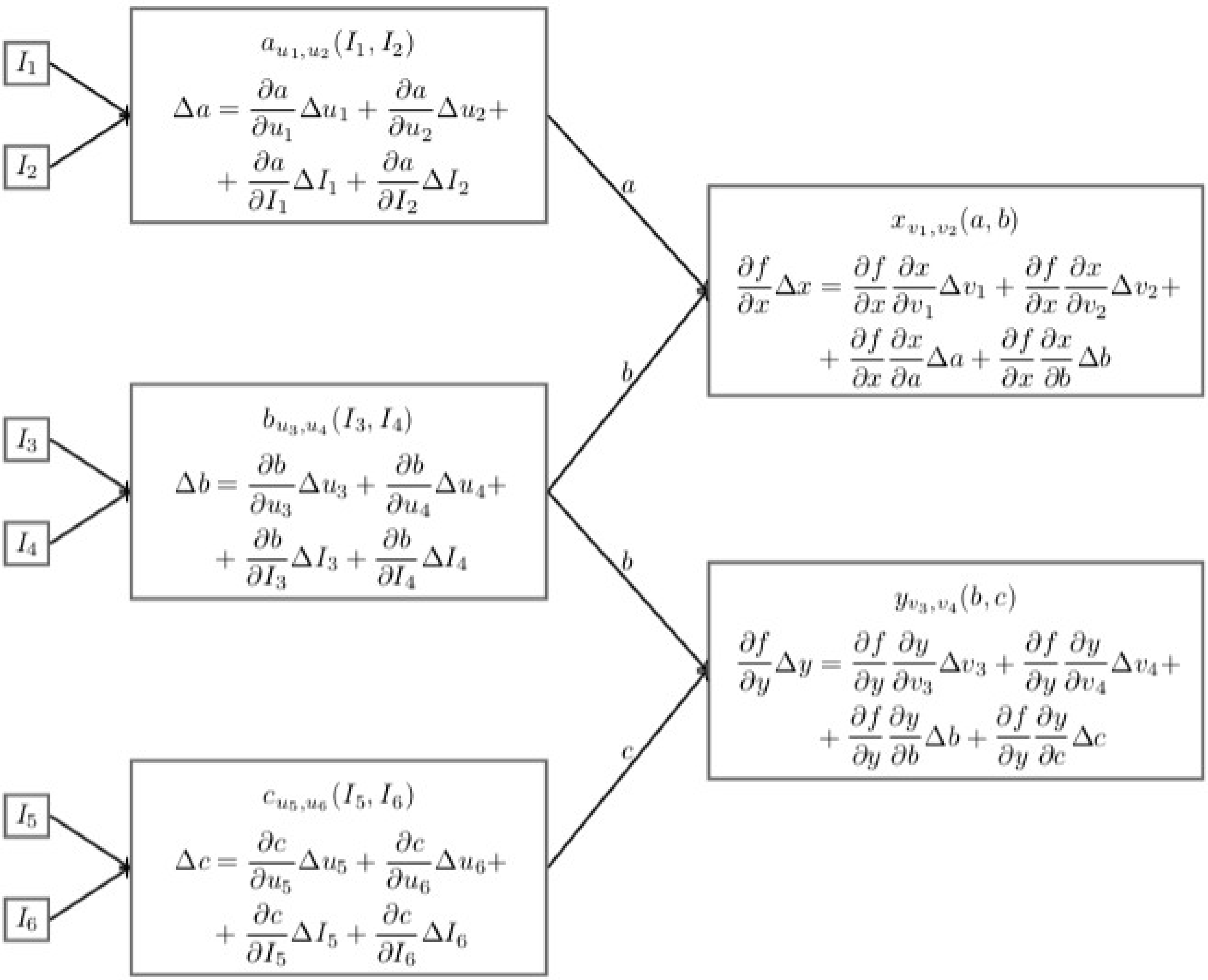
Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 +$$
$$+ \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$



Note: all derivatives are meant to be calculated at a certain point – they are numbers!

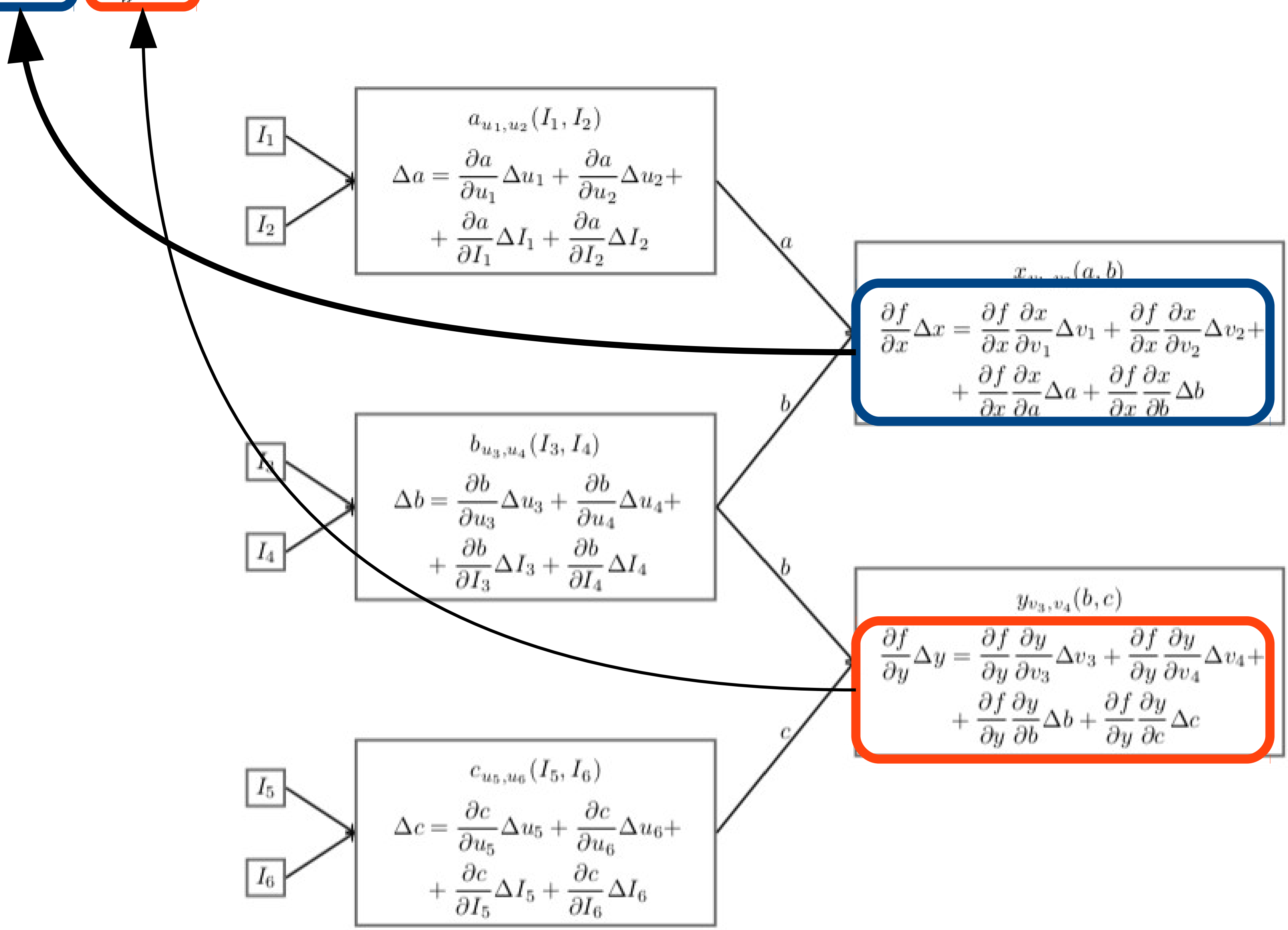
$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 +$$
$$+ \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$



Note: all derivatives are meant to be calculated at a certain point – they are numbers!

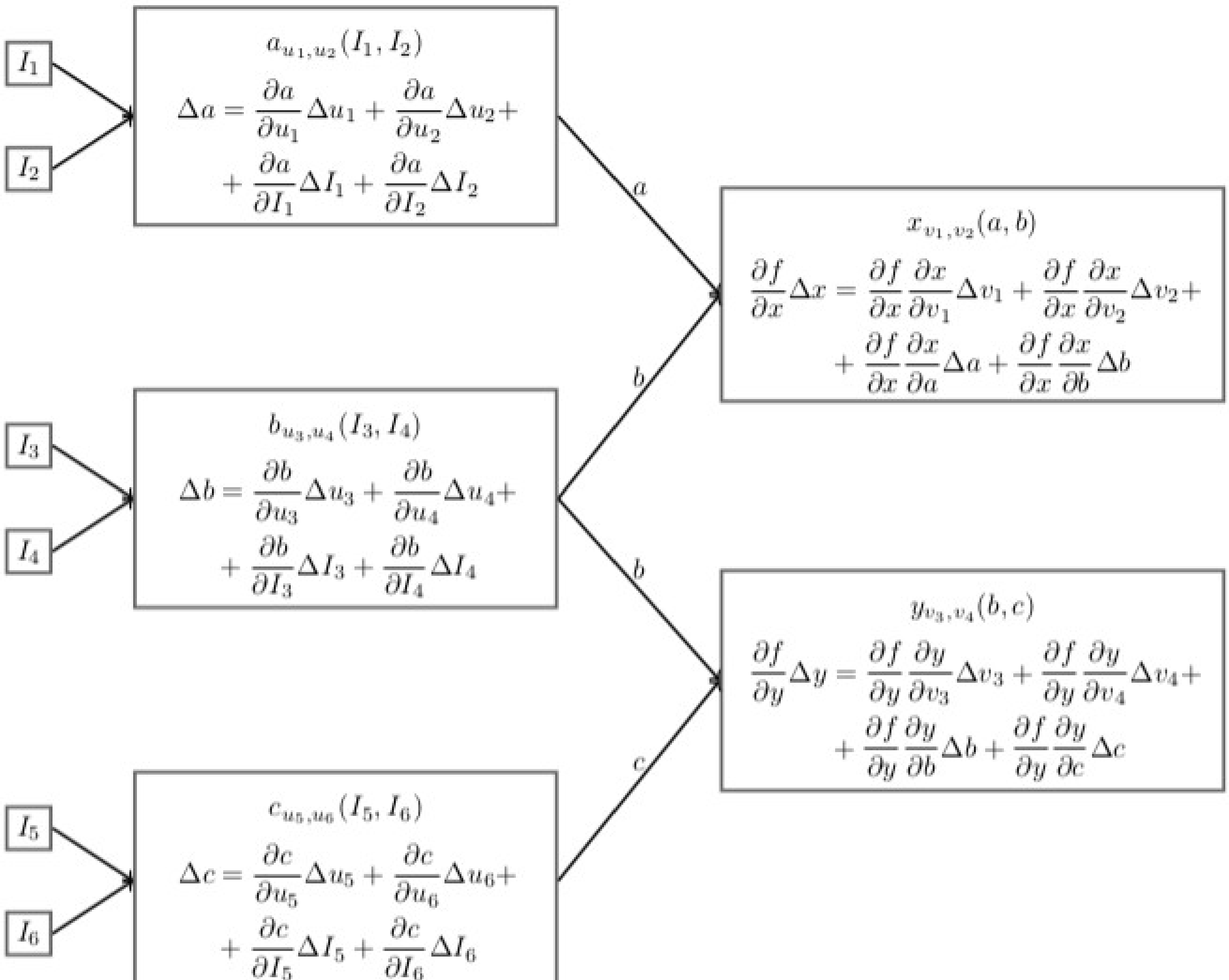
$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$

$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$



Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 +$$
$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + \frac{\partial f}{\partial x}\frac{\partial x}{\partial b}\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c$$

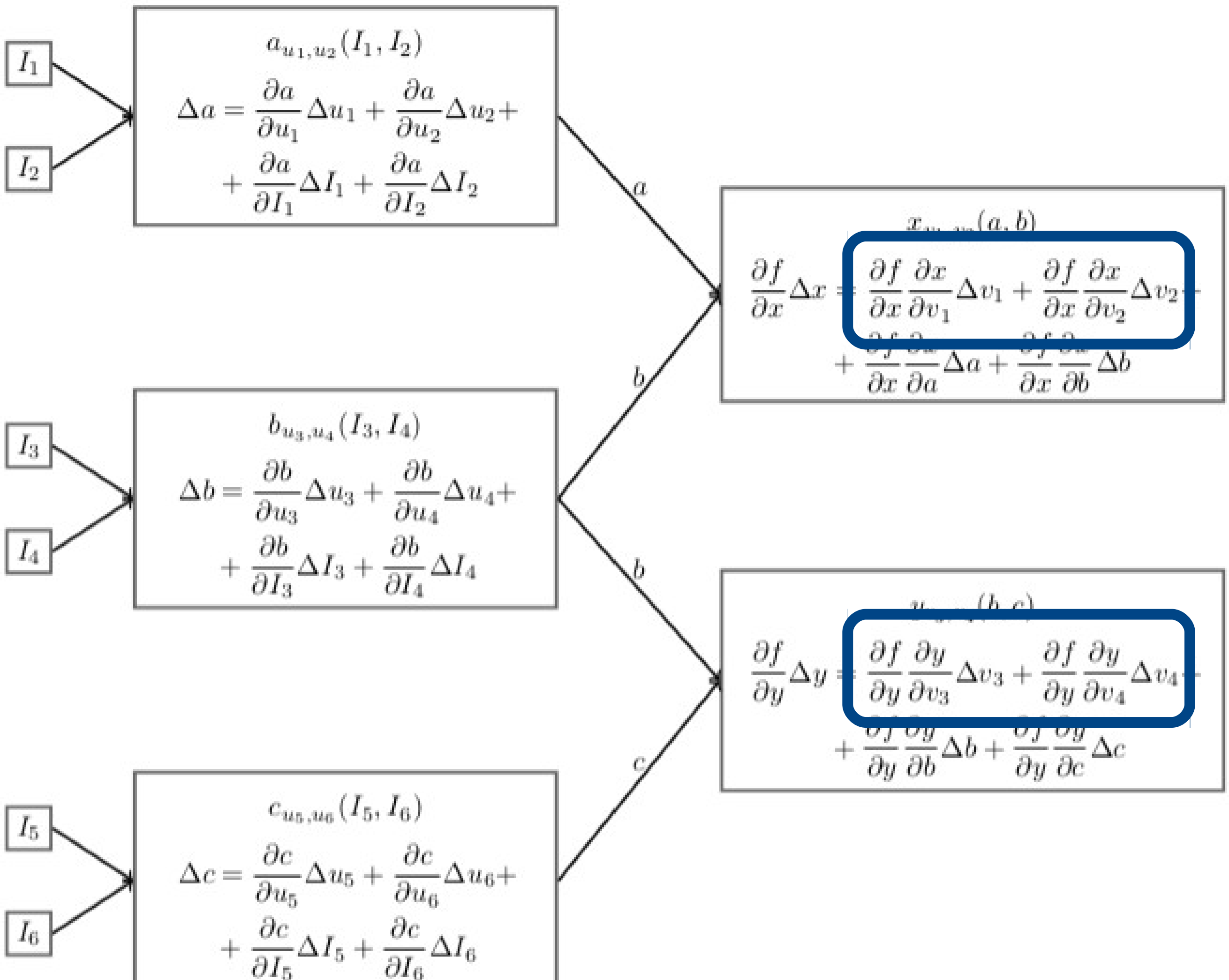


Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$

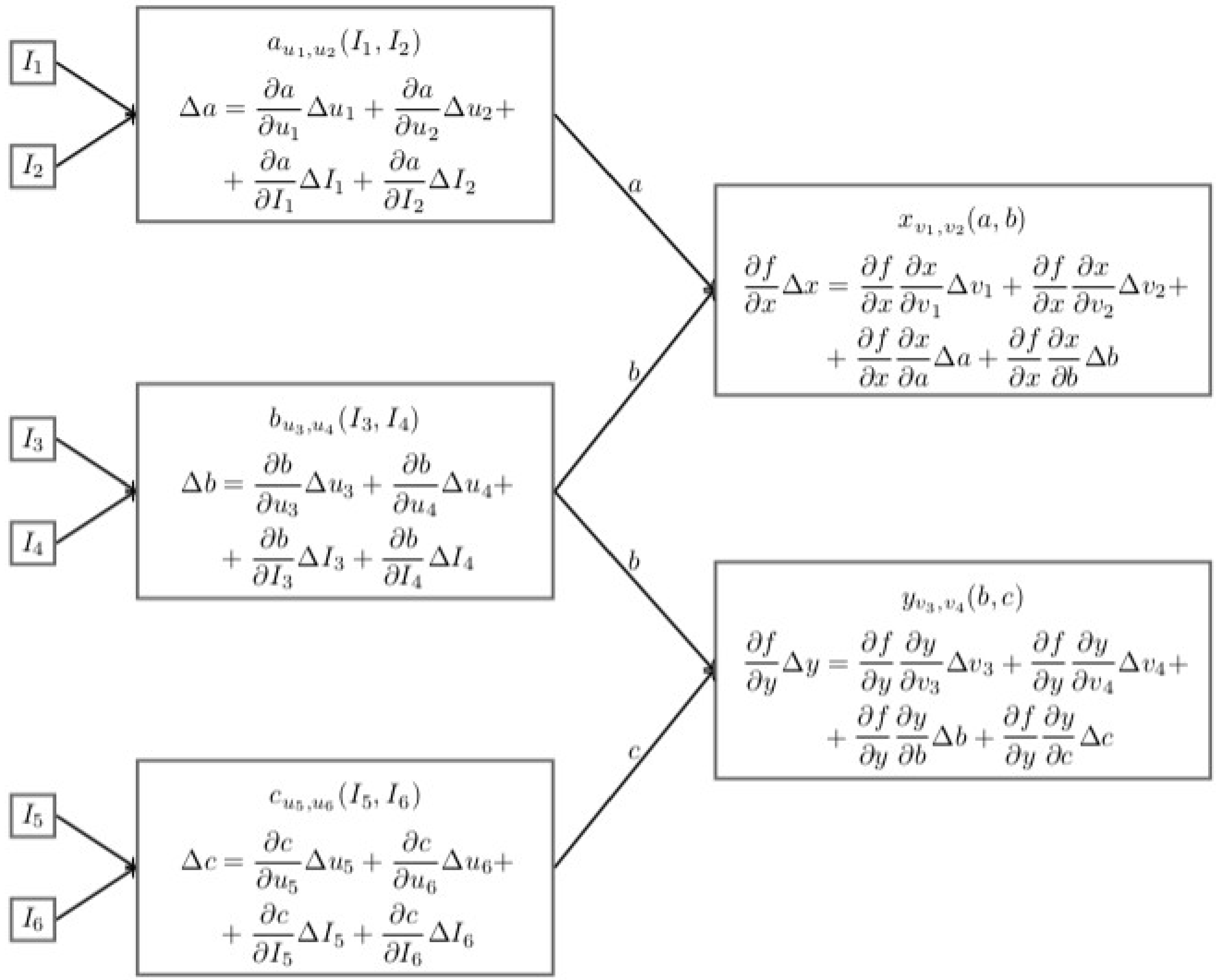
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 +$$

$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + \frac{\partial f}{\partial x}\frac{\partial x}{\partial b}\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c$$



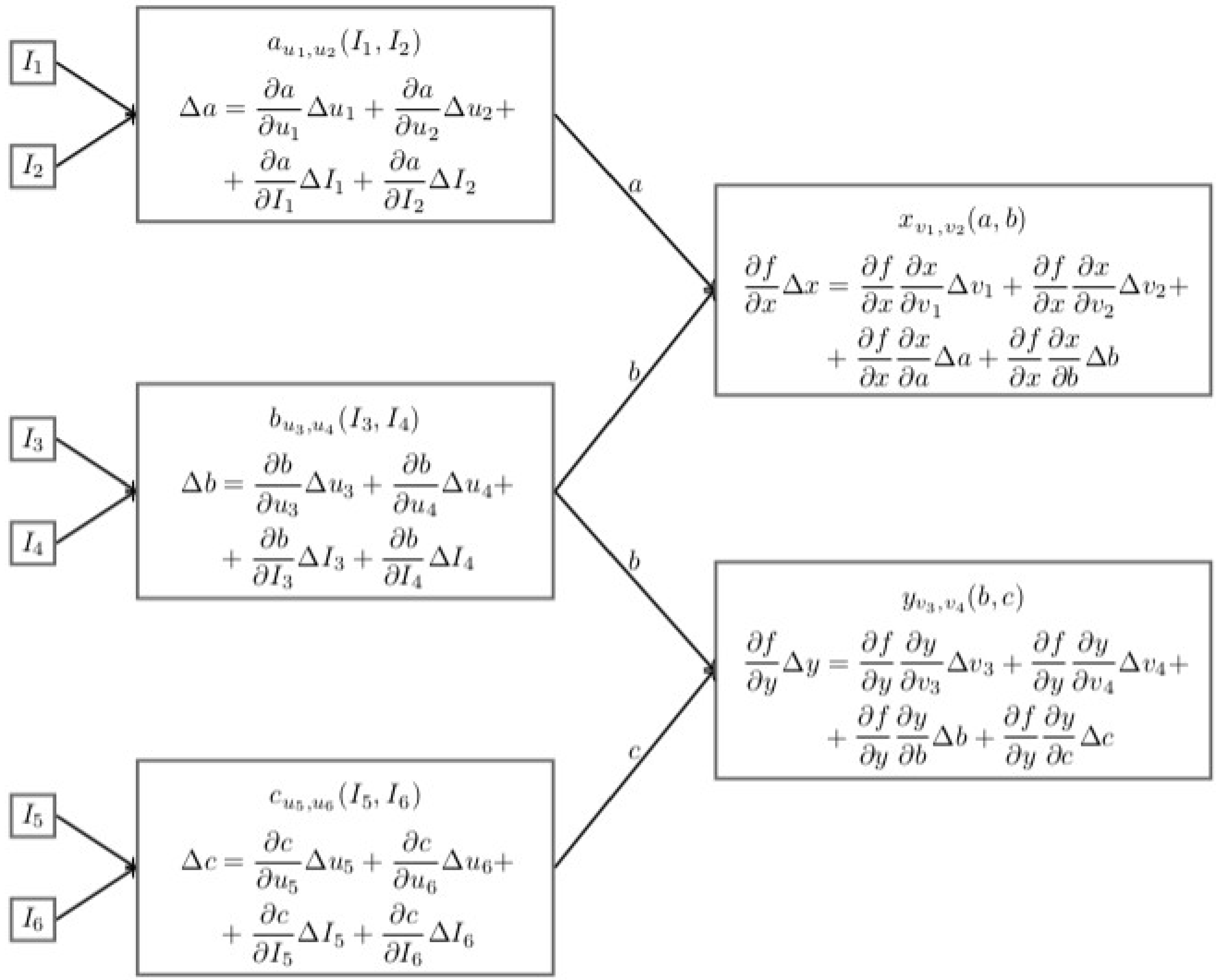
Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$
\begin{aligned}
&\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) = \\
&\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 + \\
&+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + \boxed{\frac{\partial f}{\partial x}\frac{\partial x}{\partial b}\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\Delta b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c
\end{aligned}
$$



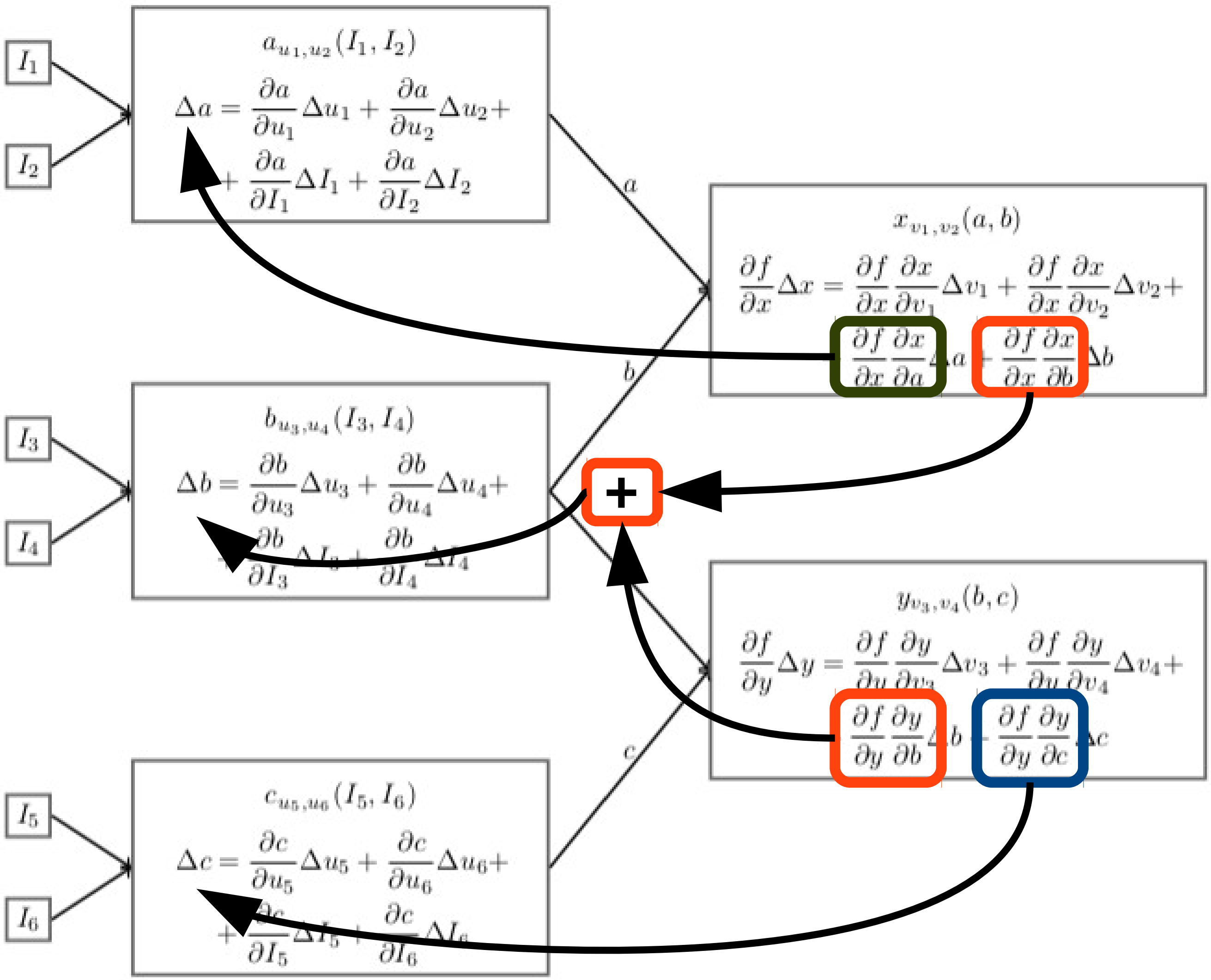
Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 +$$
$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c$$



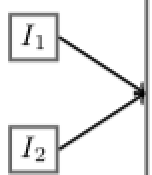
Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 +$$
$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c$$



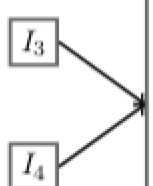
Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 +$$
$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c$$



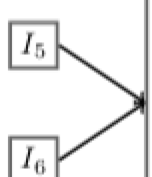
$a_{u_1,u_2}(I_1, I_2)$

$$\Delta a = \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}\Delta u_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}\Delta u_2 +$$
$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial I_1}\Delta I_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial I_2}\Delta I_2$$



$b_{u_3,u_4}(I_3, I_4)$

$$\Delta b = \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}\Delta u_3 + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}\Delta u_4 +$$
$$+ \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial I_3}\Delta I_3 + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial I_4}\Delta I_4$$



$c_{u_5,u_6}(I_5, I_6)$

$$\Delta c = \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}\Delta u_5 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}\Delta u_6 +$$
$$+ \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial I_5}\Delta I_5 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial I_6}\Delta I_6$$

Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$

$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 +$$

$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c$$

I_1, I_2 → $a_{u_1,u_2}(I_1, I_2)$

$$\Delta a = \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}\Delta u_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}\Delta u_2 +$$

$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial I_1}\Delta I_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial I_2}\Delta I_2 = 0$$

We cannot variate the input data. That would be really strange

I_3, I_4 → $b_{u_3,u_4}(I_3, I_4)$

$$\Delta b = \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}\Delta u_3 + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}\Delta u_4 +$$

$$+ \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial I_3}\Delta I_3 + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial I_4}\Delta I_4 = 0$$

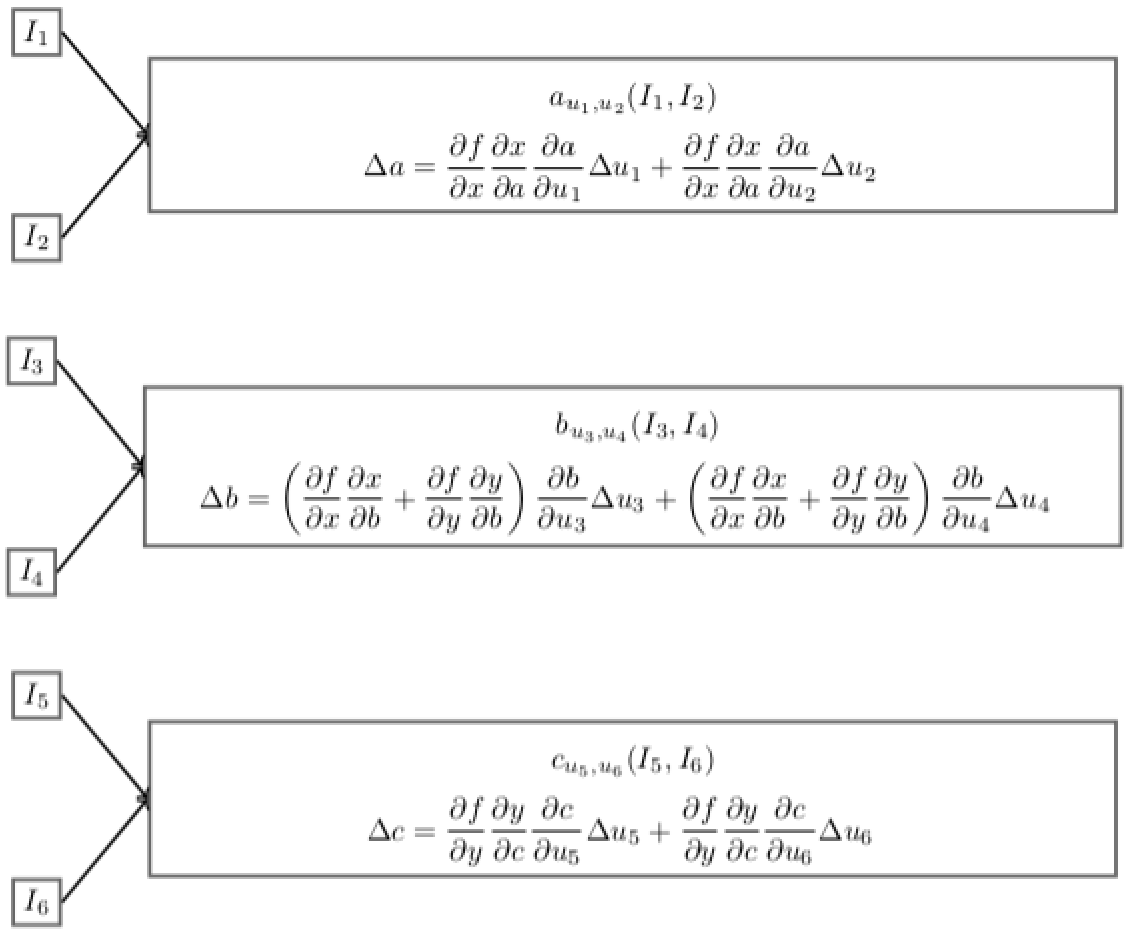
I_5, I_6 → $c_{u_5,u_6}(I_5, I_6)$

$$\Delta c = \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}\Delta u_5 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}\Delta u_6 +$$

$$+ \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial I_5}\Delta I_5 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial I_6}\Delta I_6 = 0$$

Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$
\begin{aligned}
&\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) = \\
&\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 + \\
&+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c
\end{aligned}
$$

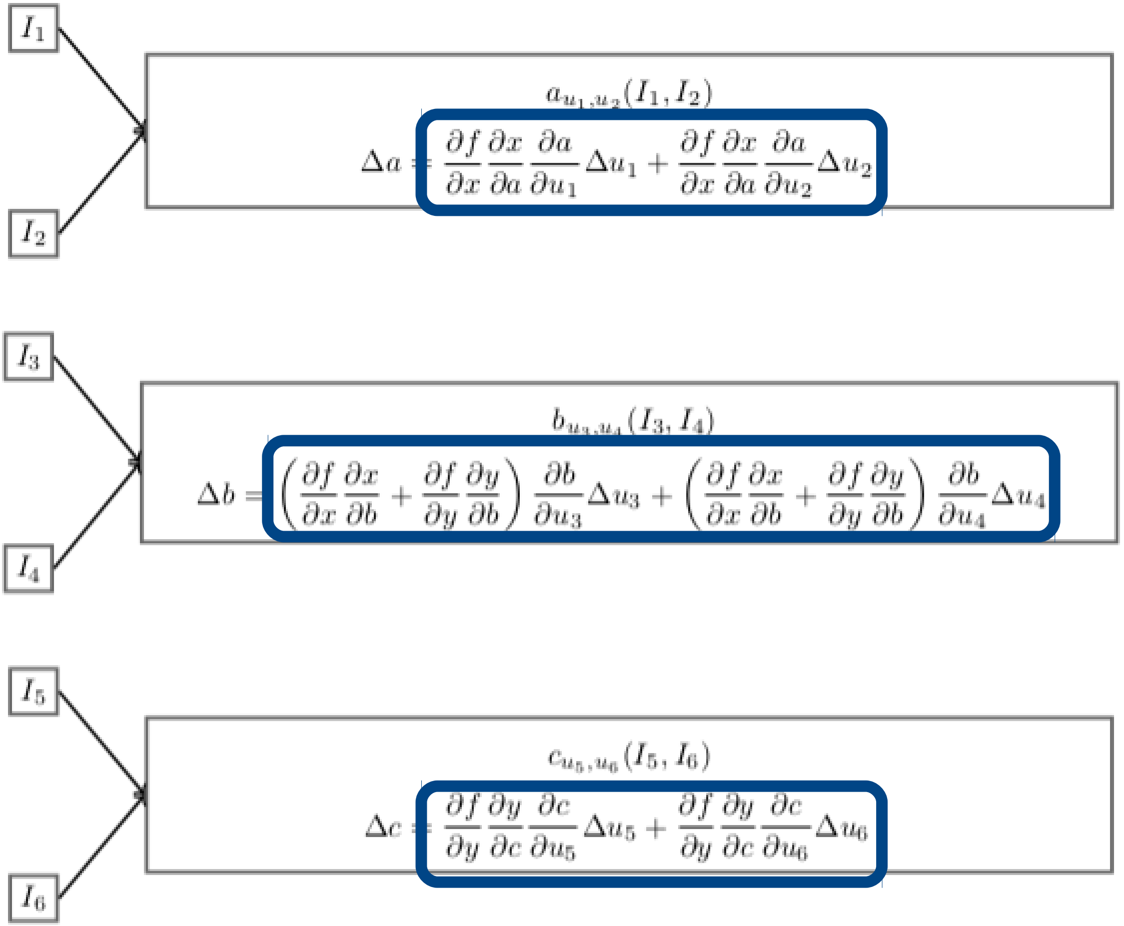


Note: all derivatives are meant to be calculated at a certain point – they are numbers!

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$

$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 +$$

$$+ \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\Delta a + + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\Delta b + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\Delta c$$



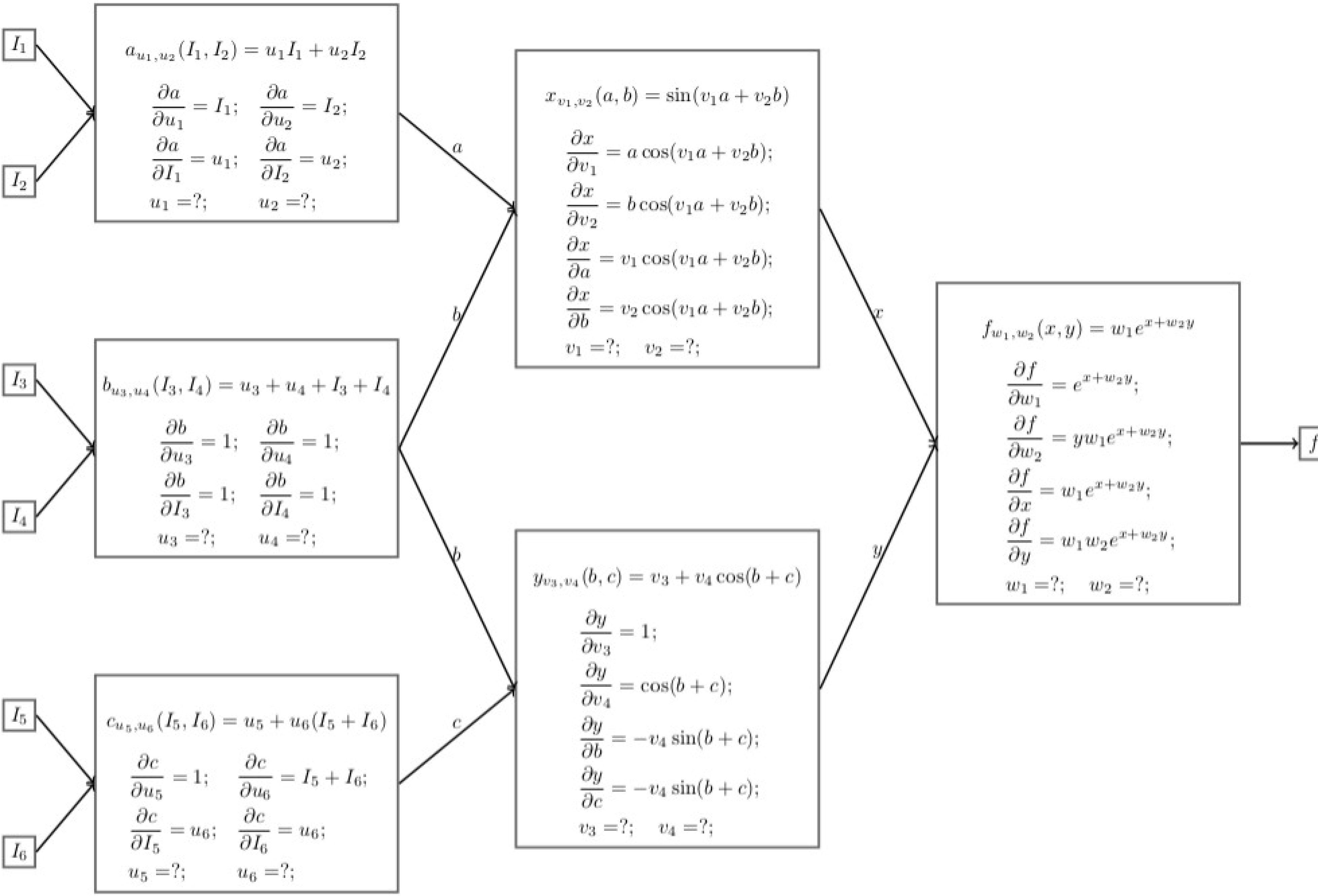
Now small variation of cost function is written in form it depends on internal parameters only

$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}\Delta v_4 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}\Delta u_1 + \frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}\Delta u_2 +$$
$$+\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}\Delta u_3 + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}\Delta u_4 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}\Delta u_5 + \frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}\Delta u_6$$

Now we have all numbers needed to make the gradient descent step!

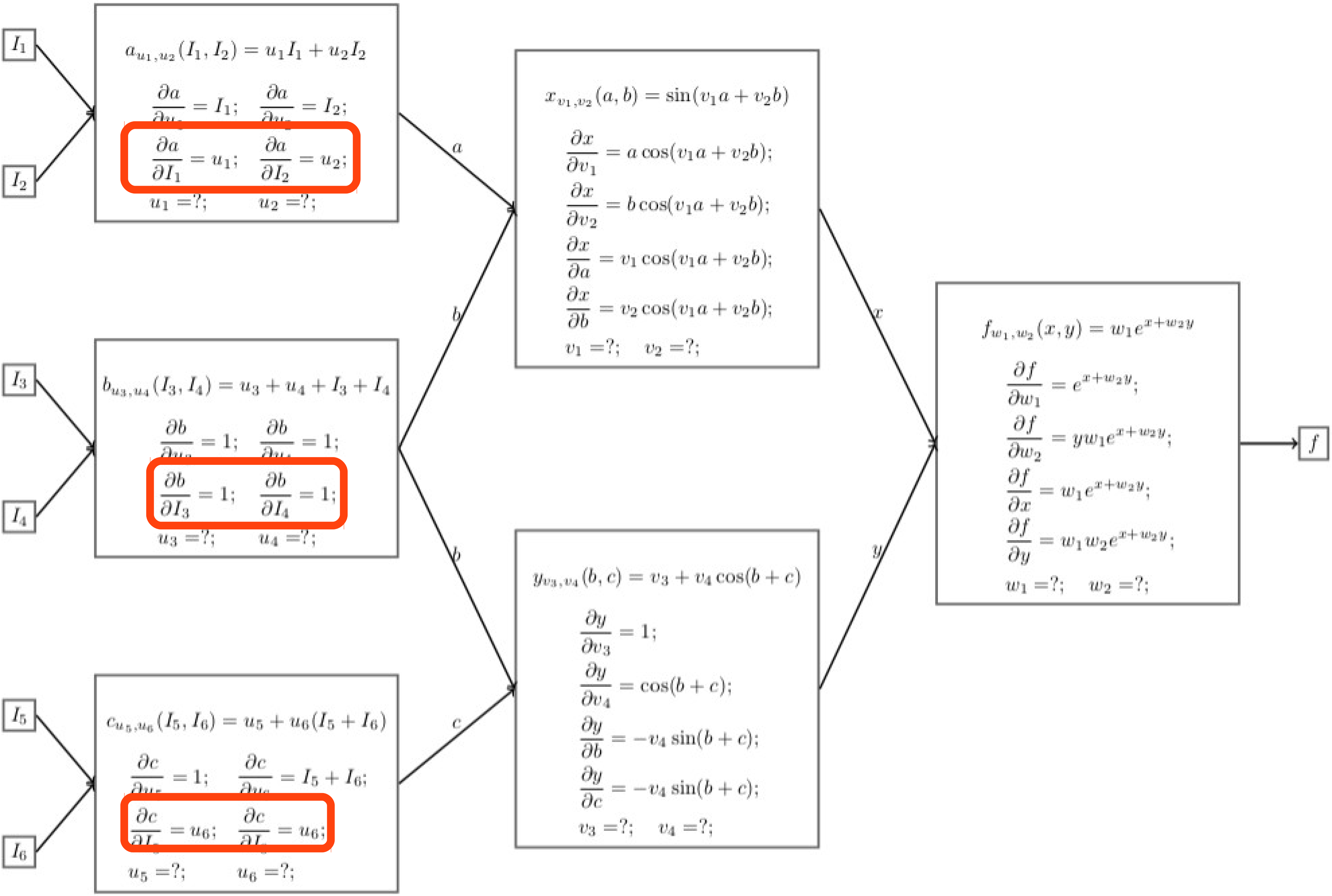
$$\Delta f(\Delta u_1, \Delta u_2, \Delta u_3, \Delta u_4, \Delta u_5, \Delta u_6, \Delta v_1, \Delta v_2, \Delta v_3, \Delta v_4, \Delta w_1, \Delta w_2) =$$
$$\frac{\partial f}{\partial w_1}\Delta w_1 + \frac{\partial f}{\partial w_2}\Delta w_2 + \frac{\partial f}{\partial v_1}\Delta v_1 + \frac{\partial f}{\partial v_2}\Delta v_2 + \frac{\partial f}{\partial v_3}\Delta v_3 + \frac{\partial f}{\partial v_4}\Delta v_4 + \frac{\partial f}{\partial u_1}\Delta u_1 + \frac{\partial f}{\partial u_2}\Delta u_2 + \frac{\partial f}{\partial u_3}\Delta u_3 + \frac{\partial f}{\partial u_4}\Delta u_4 + \frac{\partial f}{\partial u_5}\Delta u_5 + \frac{\partial f}{\partial u_6}\Delta u_6$$

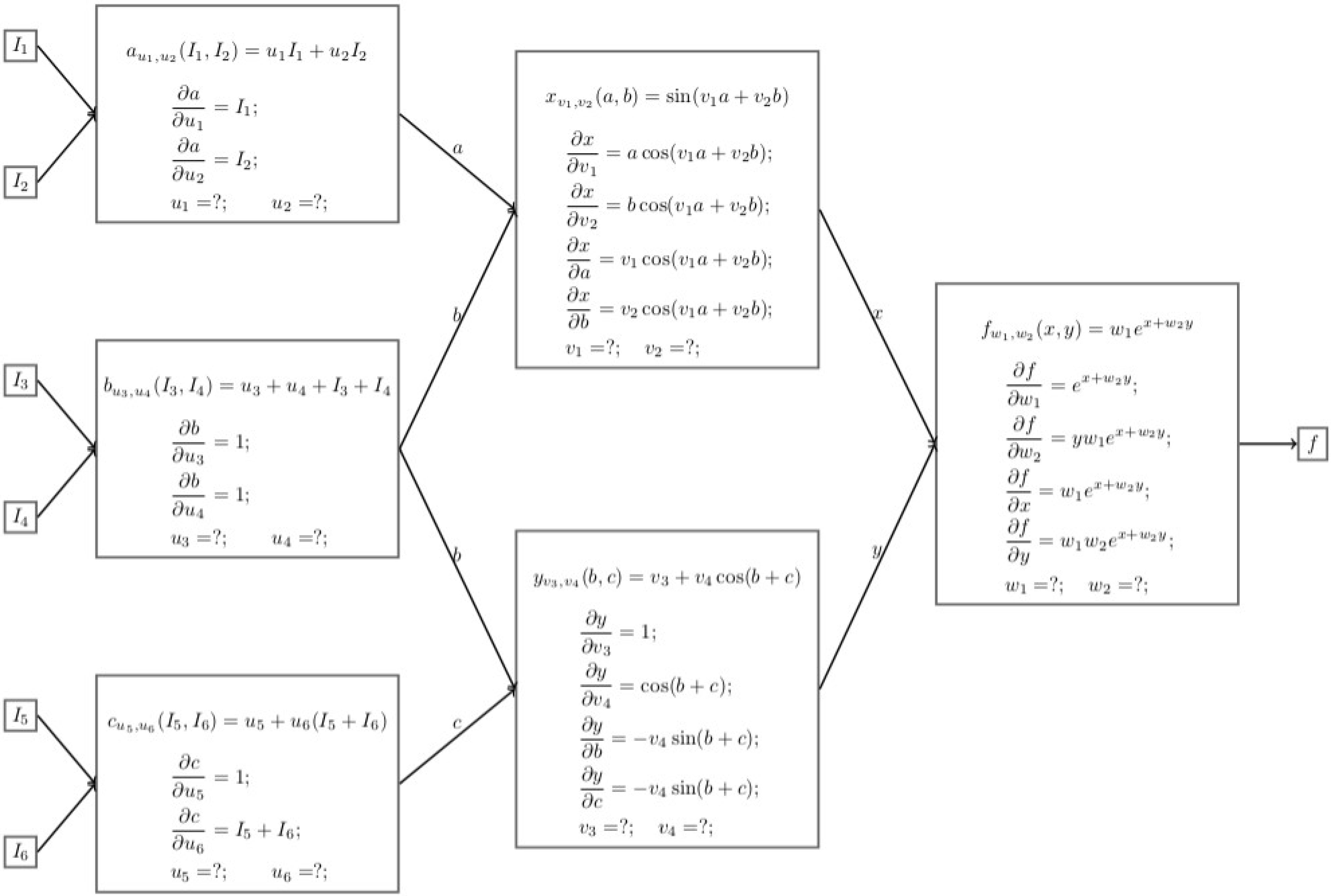
Example



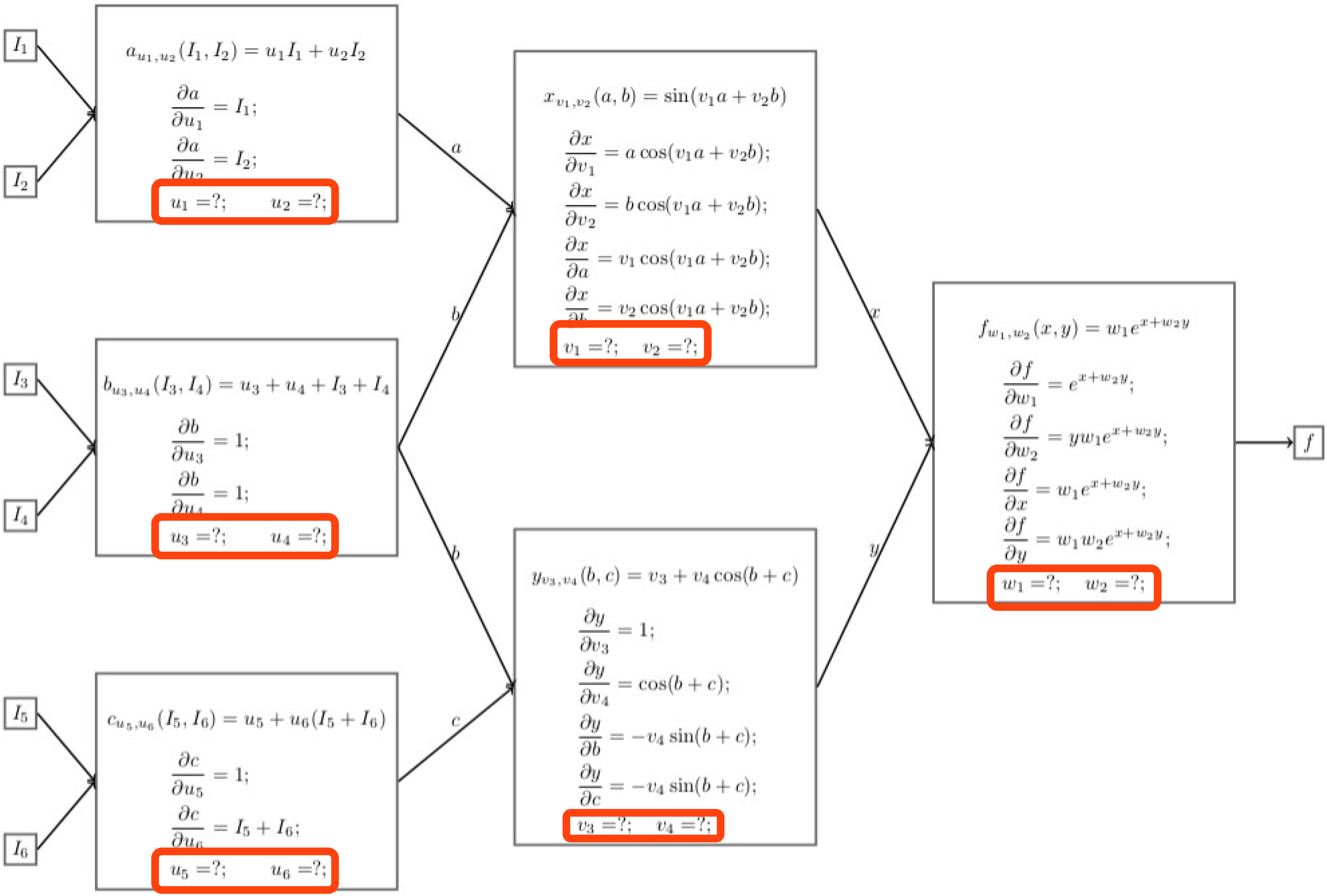
I_1
I_2
$a_{u_1,u_2}(I_1, I_2) = u_1 I_1 + u_2 I_2$
$\frac{\partial a}{\partial u_1} = I_1; \quad \frac{\partial a}{\partial u_2} = I_2;$
$\frac{\partial a}{\partial I_1} = u_1; \quad \frac{\partial a}{\partial I_2} = u_2;$
$u_1 =?; \quad u_2 =?;$
a
$x_{v_1,v_2}(a, b) = \sin(v_1 a + v_2 b)$
$\frac{\partial x}{\partial v_1} = a\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial v_2} = b\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial a} = v_1 \cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial b} = v_2 \cos(v_1 a + v_2 b);$
$v_1 =?; \quad v_2 =?;$
b
I_3
I_4
$b_{u_3,u_4}(I_3, I_4) = u_3 + u_4 + I_3 + I_4$
$\frac{\partial b}{\partial u_3} = 1; \quad \frac{\partial b}{\partial u_4} = 1;$
$\frac{\partial b}{\partial I_3} = 1; \quad \frac{\partial b}{\partial I_4} = 1;$
$u_3 =?; \quad u_4 =?;$
x
$f_{w_1,w_2}(x, y) = w_1 e^{x+w_2 y}$
$\frac{\partial f}{\partial w_1} = e^{x+w_2 y};$
$\frac{\partial f}{\partial w_2} = y w_1 e^{x+w_2 y};$
$\frac{\partial f}{\partial x} = w_1 e^{x+w_2 y};$
$\frac{\partial f}{\partial y} = w_1 w_2 e^{x+w_2 y};$
$w_1 =?; \quad w_2 =?;$
f
b
y
$y_{v_3,v_4}(b, c) = v_3 + v_4 \cos(b + c)$
$\frac{\partial y}{\partial v_3} = 1;$
$\frac{\partial y}{\partial v_4} = \cos(b + c);$
$\frac{\partial y}{\partial b} = -v_4 \sin(b + c);$
$\frac{\partial y}{\partial c} = -v_4 \sin(b + c);$
$v_3 =?; \quad v_4 =?;$
c
I_5
I_6
$c_{u_5,u_6}(I_5, I_6) = u_5 + u_6(I_5 + I_6)$
$\frac{\partial c}{\partial u_5} = 1; \quad \frac{\partial c}{\partial u_6} = I_5 + I_6;$
$\frac{\partial c}{\partial I_5} = u_6; \quad \frac{\partial c}{\partial I_6} = u_6;$
$u_5 =?; \quad u_6 =?;$

Remove unnecessary derivatives

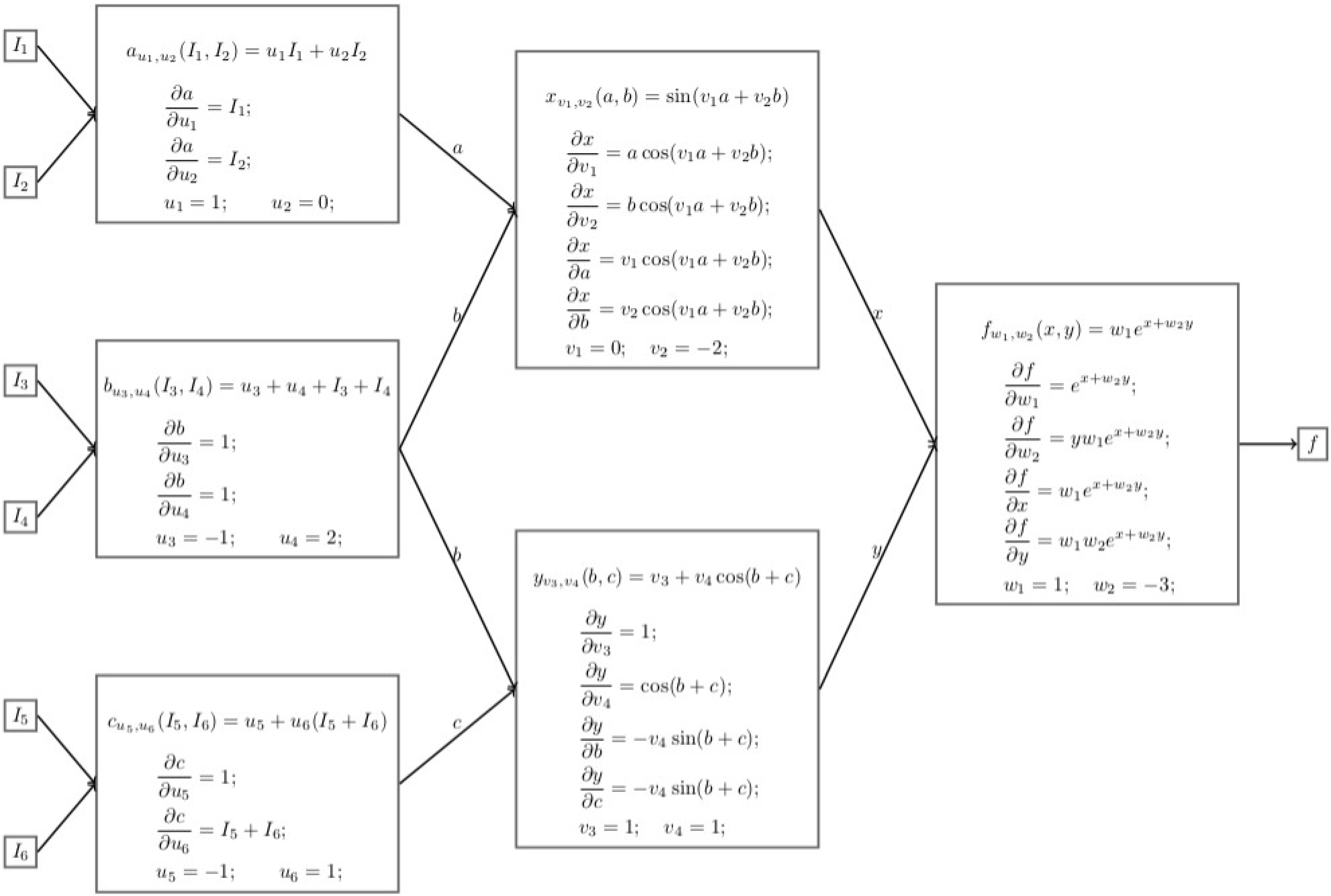



I_1
I_2
I_3
I_4
I_5
I_6
$a_{u_1,u_2}(I_1, I_2) = u_1 I_1 + u_2 I_2$
$\frac{\partial a}{\partial u_1} = I_1;$
$\frac{\partial a}{\partial u_2} = I_2;$
$u_1 = ?; \quad u_2 = ?;$
$b_{u_3,u_4}(I_3, I_4) = u_3 + u_4 + I_3 + I_4$
$\frac{\partial b}{\partial u_3} = 1;$
$\frac{\partial b}{\partial u_4} = 1;$
$u_3 = ?; \quad u_4 = ?;$
$c_{u_5,u_6}(I_5, I_6) = u_5 + u_6(I_5 + I_6)$
$\frac{\partial c}{\partial u_5} = 1;$
$\frac{\partial c}{\partial u_6} = I_5 + I_6;$
$u_5 = ?; \quad u_6 = ?;$
a
b
b
c
$x_{v_1,v_2}(a, b) = \sin(v_1 a + v_2 b)$
$\frac{\partial x}{\partial v_1} = a\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial v_2} = b\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial a} = v_1\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial b} = v_2\cos(v_1 a + v_2 b);$
$v_1 = ?; \quad v_2 = ?;$
$y_{v_3,v_4}(b, c) = v_3 + v_4\cos(b + c)$
$\frac{\partial y}{\partial v_3} = 1;$
$\frac{\partial y}{\partial v_4} = \cos(b + c);$
$\frac{\partial y}{\partial b} = -v_4\sin(b + c);$
$\frac{\partial y}{\partial c} = -v_4\sin(b + c);$
$v_3 = ?; \quad v_4 = ?;$
x
y
$f_{w_1,w_2}(x, y) = w_1 e^{x + w_2 y}$
$\frac{\partial f}{\partial w_1} = e^{x + w_2 y};$
$\frac{\partial f}{\partial w_2} = y w_1 e^{x + w_2 y};$
$\frac{\partial f}{\partial x} = w_1 e^{x + w_2 y};$
$\frac{\partial f}{\partial y} = w_1 w_2 e^{x + w_2 y};$
$w_1 = ?; \quad w_2 = ?;$
f

Initialize parameters

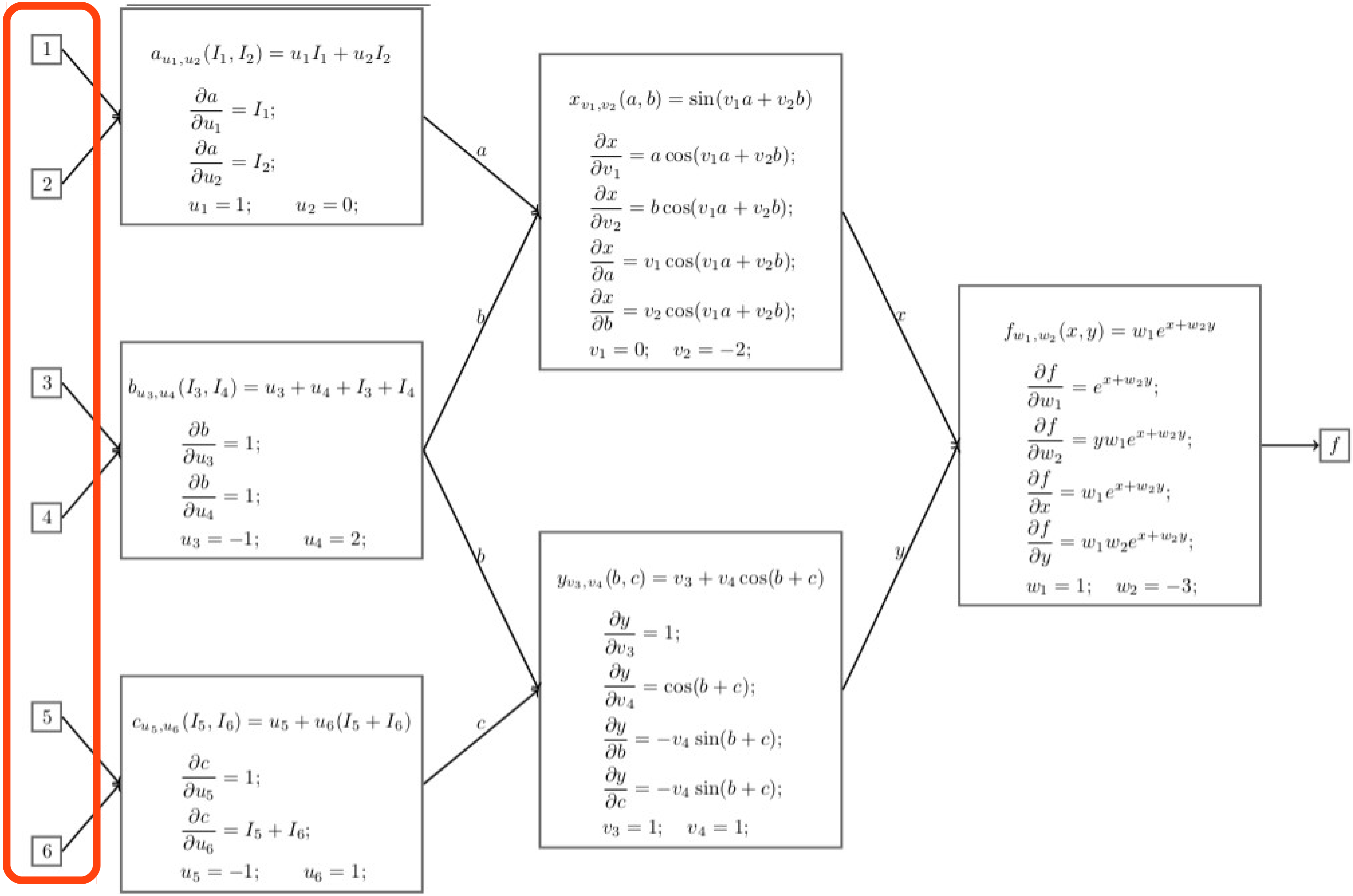


Forward



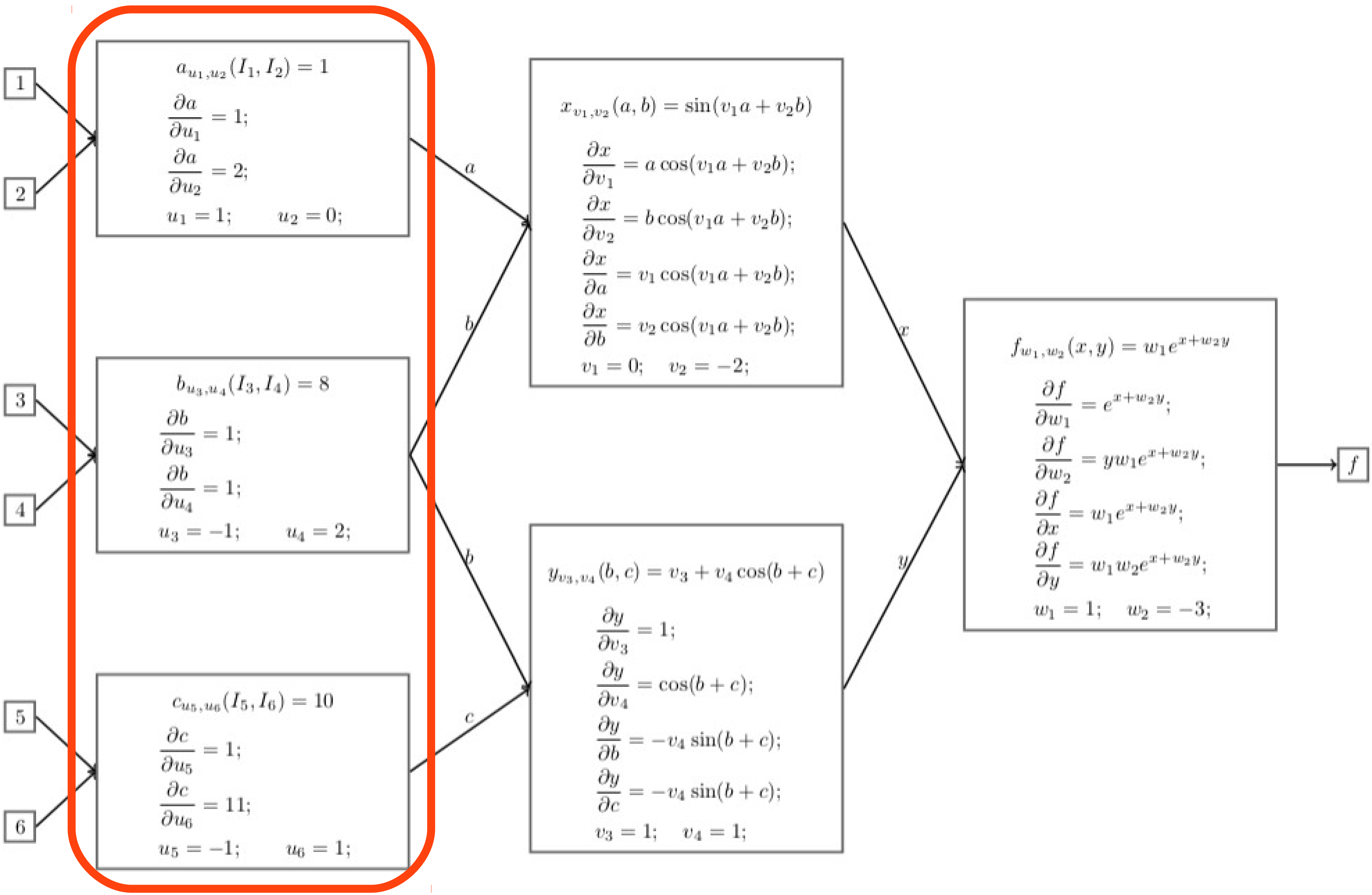
I_1
I_2
$a_{u_1,u_2}(I_1, I_2) = u_1 I_1 + u_2 I_2$
$\frac{\partial a}{\partial u_1} = I_1;$
$\frac{\partial a}{\partial u_2} = I_2;$
$u_1 = 1; \quad u_2 = 0;$
I_3
I_4
$b_{u_3,u_4}(I_3, I_4) = u_3 + u_4 + I_3 + I_4$
$\frac{\partial b}{\partial u_3} = 1;$
$\frac{\partial b}{\partial u_4} = 1;$
$u_3 = -1; \quad u_4 = 2;$
I_5
I_6
$c_{u_5,u_6}(I_5, I_6) = u_5 + u_6(I_5 + I_6)$
$\frac{\partial c}{\partial u_5} = 1;$
$\frac{\partial c}{\partial u_6} = I_5 + I_6;$
$u_5 = -1; \quad u_6 = 1;$
a
b
b
c
$x_{v_1,v_2}(a, b) = \sin(v_1 a + v_2 b)$
$\frac{\partial x}{\partial v_1} = a\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial v_2} = b\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial a} = v_1 \cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial b} = v_2 \cos(v_1 a + v_2 b);$
$v_1 = 0; \quad v_2 = -2;$
$y_{v_3,v_4}(b, c) = v_3 + v_4 \cos(b + c)$
$\frac{\partial y}{\partial v_3} = 1;$
$\frac{\partial y}{\partial v_4} = \cos(b + c);$
$\frac{\partial y}{\partial b} = -v_4 \sin(b + c);$
$\frac{\partial y}{\partial c} = -v_4 \sin(b + c);$
$v_3 = 1; \quad v_4 = 1;$
x
y
$f_{w_1,w_2}(x, y) = w_1 e^{x + w_2 y}$
$\frac{\partial f}{\partial w_1} = e^{x + w_2 y};$
$\frac{\partial f}{\partial w_2} = y w_1 e^{x + w_2 y};$
$\frac{\partial f}{\partial x} = w_1 e^{x + w_2 y};$
$\frac{\partial f}{\partial y} = w_1 w_2 e^{x + w_2 y};$
$w_1 = 1; \quad w_2 = -3;$
f

Forward



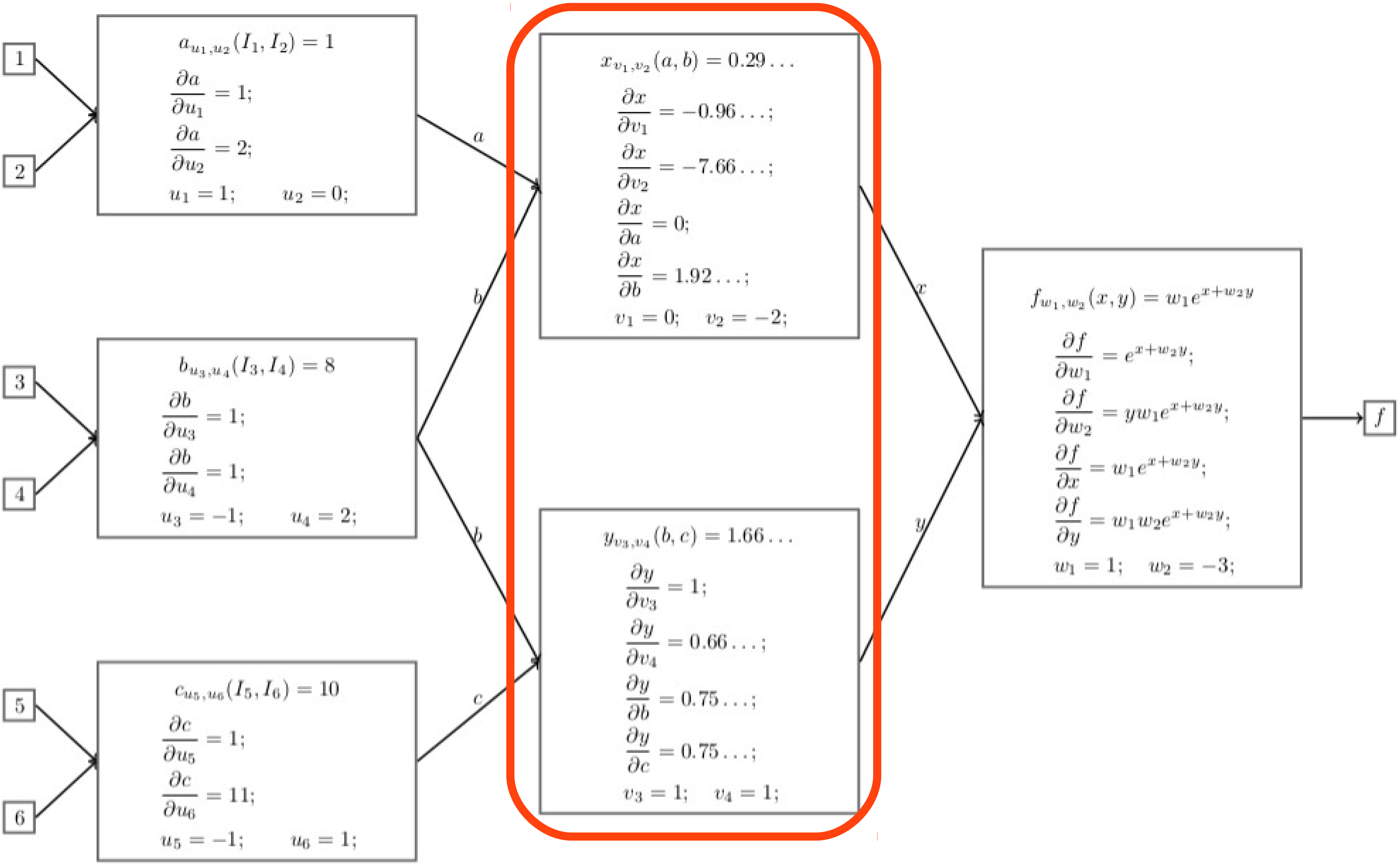
1
2
3
4
5
6
$a_{u_1,u_2}(I_1, I_2) = u_1 I_1 + u_2 I_2$
$\frac{\partial a}{\partial u_1} = I_1;$
$\frac{\partial a}{\partial u_2} = I_2;$
$u_1 = 1; \quad u_2 = 0;$
$b_{u_3,u_4}(I_3, I_4) = u_3 + u_4 + I_3 + I_4$
$\frac{\partial b}{\partial u_3} = 1;$
$\frac{\partial b}{\partial u_4} = 1;$
$u_3 = -1; \quad u_4 = 2;$
$c_{u_5,u_6}(I_5, I_6) = u_5 + u_6(I_5 + I_6)$
$\frac{\partial c}{\partial u_5} = 1;$
$\frac{\partial c}{\partial u_6} = I_5 + I_6;$
$u_5 = -1; \quad u_6 = 1;$
a
b
b
c
$x_{v_1,v_2}(a, b) = \sin(v_1 a + v_2 b)$
$\frac{\partial x}{\partial v_1} = a\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial v_2} = b\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial a} = v_1\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial b} = v_2\cos(v_1 a + v_2 b);$
$v_1 = 0; \quad v_2 = -2;$
$y_{v_3,v_4}(b, c) = v_3 + v_4\cos(b + c)$
$\frac{\partial y}{\partial v_3} = 1;$
$\frac{\partial y}{\partial v_4} = \cos(b + c);$
$\frac{\partial y}{\partial b} = -v_4\sin(b + c);$
$\frac{\partial y}{\partial c} = -v_4\sin(b + c);$
$v_3 = 1; \quad v_4 = 1;$
x
y
$f_{w_1,w_2}(x, y) = w_1 e^{x + w_2 y}$
$\frac{\partial f}{\partial w_1} = e^{x + w_2 y};$
$\frac{\partial f}{\partial w_2} = y w_1 e^{x + w_2 y};$
$\frac{\partial f}{\partial x} = w_1 e^{x + w_2 y};$
$\frac{\partial f}{\partial y} = w_1 w_2 e^{x + w_2 y};$
$w_1 = 1; \quad w_2 = -3;$
f

Forward



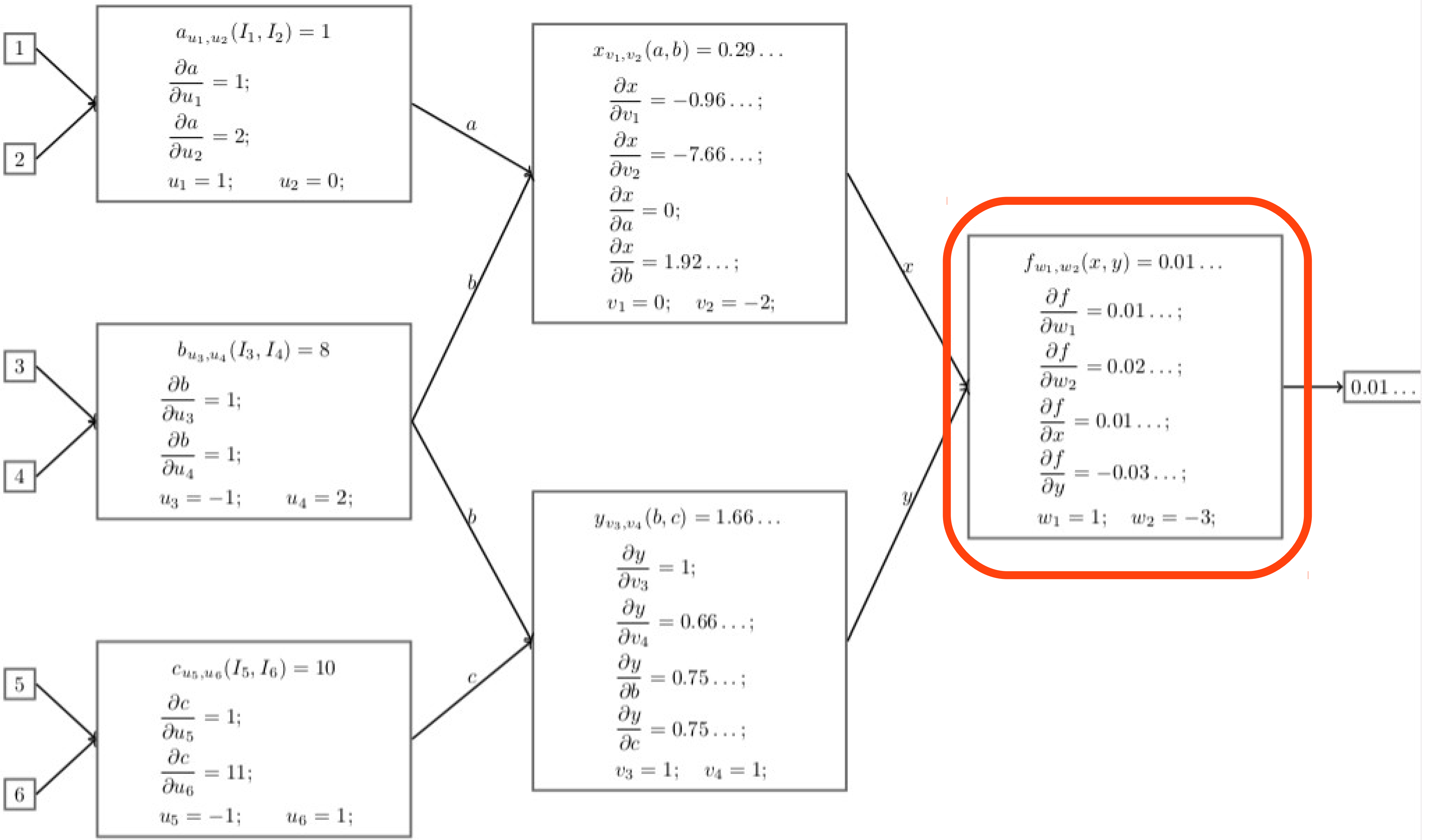
1
2
3
4
5
6
$a_{u_1,u_2}(I_1, I_2) = 1$
$\frac{\partial a}{\partial u_1} = 1;$
$\frac{\partial a}{\partial u_2} = 2;$
$u_1 = 1; \quad u_2 = 0;$
$b_{u_3,u_4}(I_3, I_4) = 8$
$\frac{\partial b}{\partial u_3} = 1;$
$\frac{\partial b}{\partial u_4} = 1;$
$u_3 = -1; \quad u_4 = 2;$
$c_{u_5,u_6}(I_5, I_6) = 10$
$\frac{\partial c}{\partial u_5} = 1;$
$\frac{\partial c}{\partial u_6} = 11;$
$u_5 = -1; \quad u_6 = 1;$
a
b
b
c
$x_{v_1,v_2}(a, b) = \sin(v_1 a + v_2 b)$
$\frac{\partial x}{\partial v_1} = a\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial v_2} = b\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial a} = v_1\cos(v_1 a + v_2 b);$
$\frac{\partial x}{\partial b} = v_2\cos(v_1 a + v_2 b);$
$v_1 = 0; \quad v_2 = -2;$
$y_{v_3,v_4}(b, c) = v_3 + v_4\cos(b + c)$
$\frac{\partial y}{\partial v_3} = 1;$
$\frac{\partial y}{\partial v_4} = \cos(b + c);$
$\frac{\partial y}{\partial b} = -v_4\sin(b + c);$
$\frac{\partial y}{\partial c} = -v_4\sin(b + c);$
$v_3 = 1; \quad v_4 = 1;$
x
y
$f_{w_1,w_2}(x, y) = w_1 e^{x+w_2 y}$
$\frac{\partial f}{\partial w_1} = e^{x+w_2 y};$
$\frac{\partial f}{\partial w_2} = y w_1 e^{x+w_2 y};$
$\frac{\partial f}{\partial x} = w_1 e^{x+w_2 y};$
$\frac{\partial f}{\partial y} = w_1 w_2 e^{x+w_2 y};$
$w_1 = 1; \quad w_2 = -3;$
f

Forward



1
2
3
4
5
6
$a_{u_1,u_2}(I_1, I_2) = 1$
$\frac{\partial a}{\partial u_1} = 1;$
$\frac{\partial a}{\partial u_2} = 2;$
$u_1 = 1; \quad u_2 = 0;$
$b_{u_3,u_4}(I_3, I_4) = 8$
$\frac{\partial b}{\partial u_3} = 1;$
$\frac{\partial b}{\partial u_4} = 1;$
$u_3 = -1; \quad u_4 = 2;$
$c_{u_5,u_6}(I_5, I_6) = 10$
$\frac{\partial c}{\partial u_5} = 1;$
$\frac{\partial c}{\partial u_6} = 11;$
$u_5 = -1; \quad u_6 = 1;$
a
b
b
c
$x_{v_1,v_2}(a, b) = 0.29\ldots$
$\frac{\partial x}{\partial v_1} = -0.96\ldots;$
$\frac{\partial x}{\partial v_2} = -7.66\ldots;$
$\frac{\partial x}{\partial a} = 0;$
$\frac{\partial x}{\partial b} = 1.92\ldots;$
$v_1 = 0; \quad v_2 = -2;$
$y_{v_3,v_4}(b, c) = 1.66\ldots$
$\frac{\partial y}{\partial v_3} = 1;$
$\frac{\partial y}{\partial v_4} = 0.66\ldots;$
$\frac{\partial y}{\partial b} = 0.75\ldots;$
$\frac{\partial y}{\partial c} = 0.75\ldots;$
$v_3 = 1; \quad v_4 = 1;$
x
y
$f_{w_1,w_2}(x, y) = w_1 e^{x+w_2 y}$
$\frac{\partial f}{\partial w_1} = e^{x+w_2 y};$
$\frac{\partial f}{\partial w_2} = y w_1 e^{x+w_2 y};$
$\frac{\partial f}{\partial x} = w_1 e^{x+w_2 y};$
$\frac{\partial f}{\partial y} = w_1 w_2 e^{x+w_2 y};$
$w_1 = 1; \quad w_2 = -3;$
f

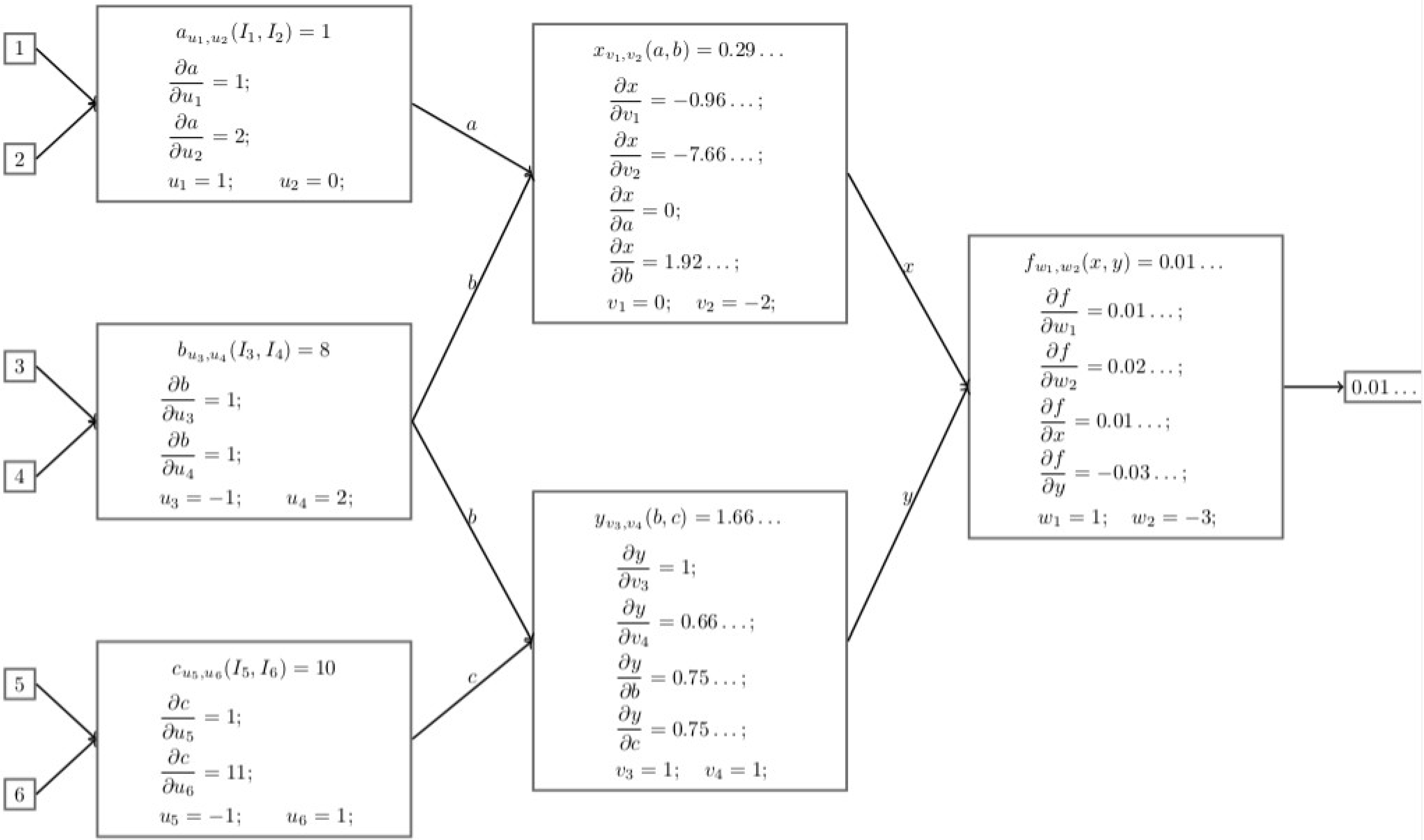
Forward



1
2
3
4
5
6
$a_{u_1,u_2}(I_1, I_2) = 1$
$\frac{\partial a}{\partial u_1} = 1;$
$\frac{\partial a}{\partial u_2} = 2;$
$u_1 = 1; \quad u_2 = 0;$
$b_{u_3,u_4}(I_3, I_4) = 8$
$\frac{\partial b}{\partial u_3} = 1;$
$\frac{\partial b}{\partial u_4} = 1;$
$u_3 = -1; \quad u_4 = 2;$
$c_{u_5,u_6}(I_5, I_6) = 10$
$\frac{\partial c}{\partial u_5} = 1;$
$\frac{\partial c}{\partial u_6} = 11;$
$u_5 = -1; \quad u_6 = 1;$
a
b
b
c
$x_{v_1,v_2}(a, b) = 0.29\ldots$
$\frac{\partial x}{\partial v_1} = -0.96\ldots;$
$\frac{\partial x}{\partial v_2} = -7.66\ldots;$
$\frac{\partial x}{\partial a} = 0;$
$\frac{\partial x}{\partial b} = 1.92\ldots;$
$v_1 = 0; \quad v_2 = -2;$
$y_{v_3,v_4}(b, c) = 1.66\ldots$
$\frac{\partial y}{\partial v_3} = 1;$
$\frac{\partial y}{\partial v_4} = 0.66\ldots;$
$\frac{\partial y}{\partial b} = 0.75\ldots;$
$\frac{\partial y}{\partial c} = 0.75\ldots;$
$v_3 = 1; \quad v_4 = 1;$
x
y
$f_{w_1,w_2}(x, y) = 0.01\ldots$
$\frac{\partial f}{\partial w_1} = 0.01\ldots;$
$\frac{\partial f}{\partial w_2} = 0.02\ldots;$
$\frac{\partial f}{\partial x} = 0.01\ldots;$
$\frac{\partial f}{\partial y} = -0.03\ldots;$
$w_1 = 1; \quad w_2 = -3;$
0.01...

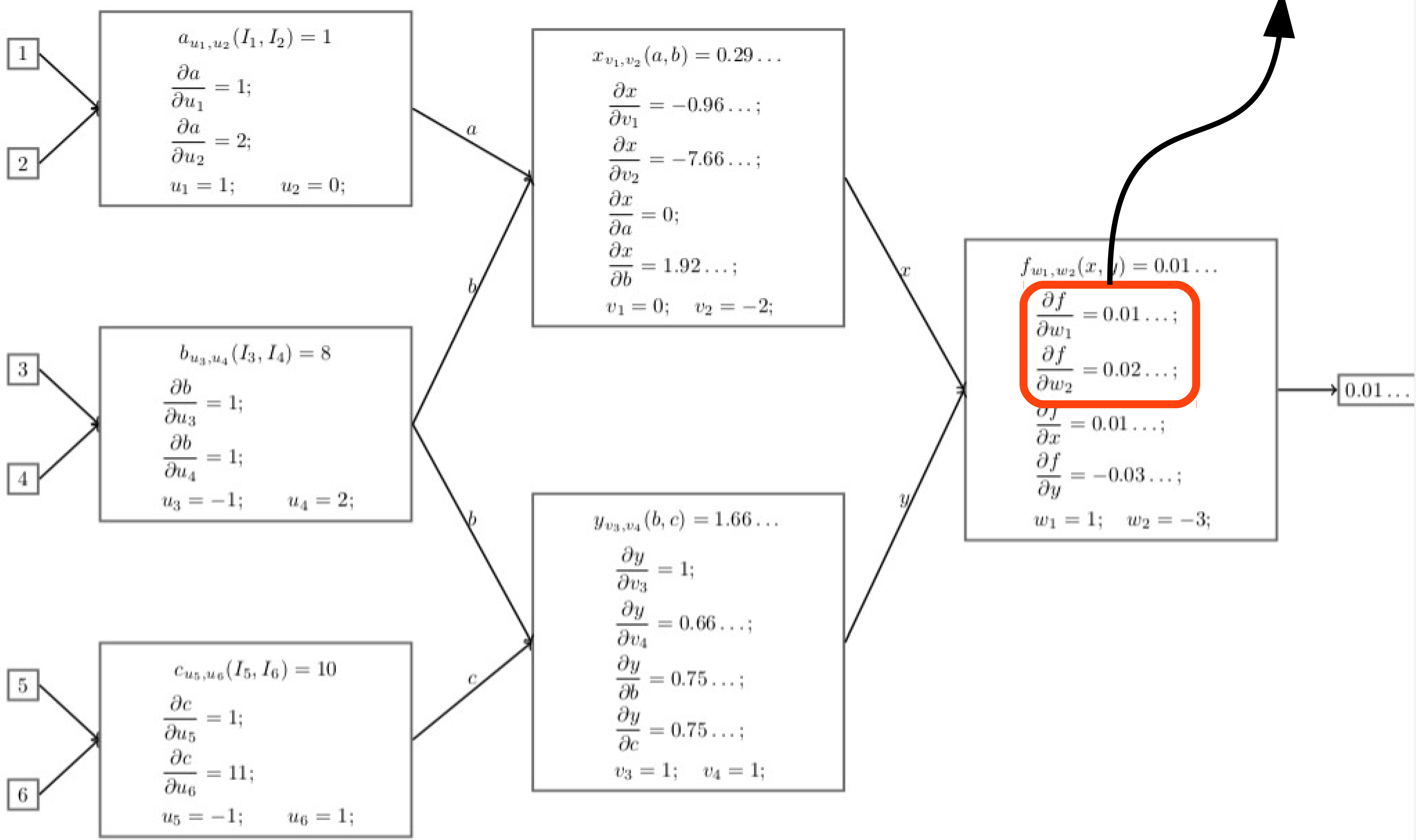
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}}_{\Delta v_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}}_{\Delta v_2}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}}_{\Delta v_3}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}}_{\Delta v_4}, \underbrace{\frac{\partial f}{\partial w_1}}_{\Delta w_1}, \underbrace{\frac{\partial f}{\partial w_2}}_{\Delta w_2} \right\}$$



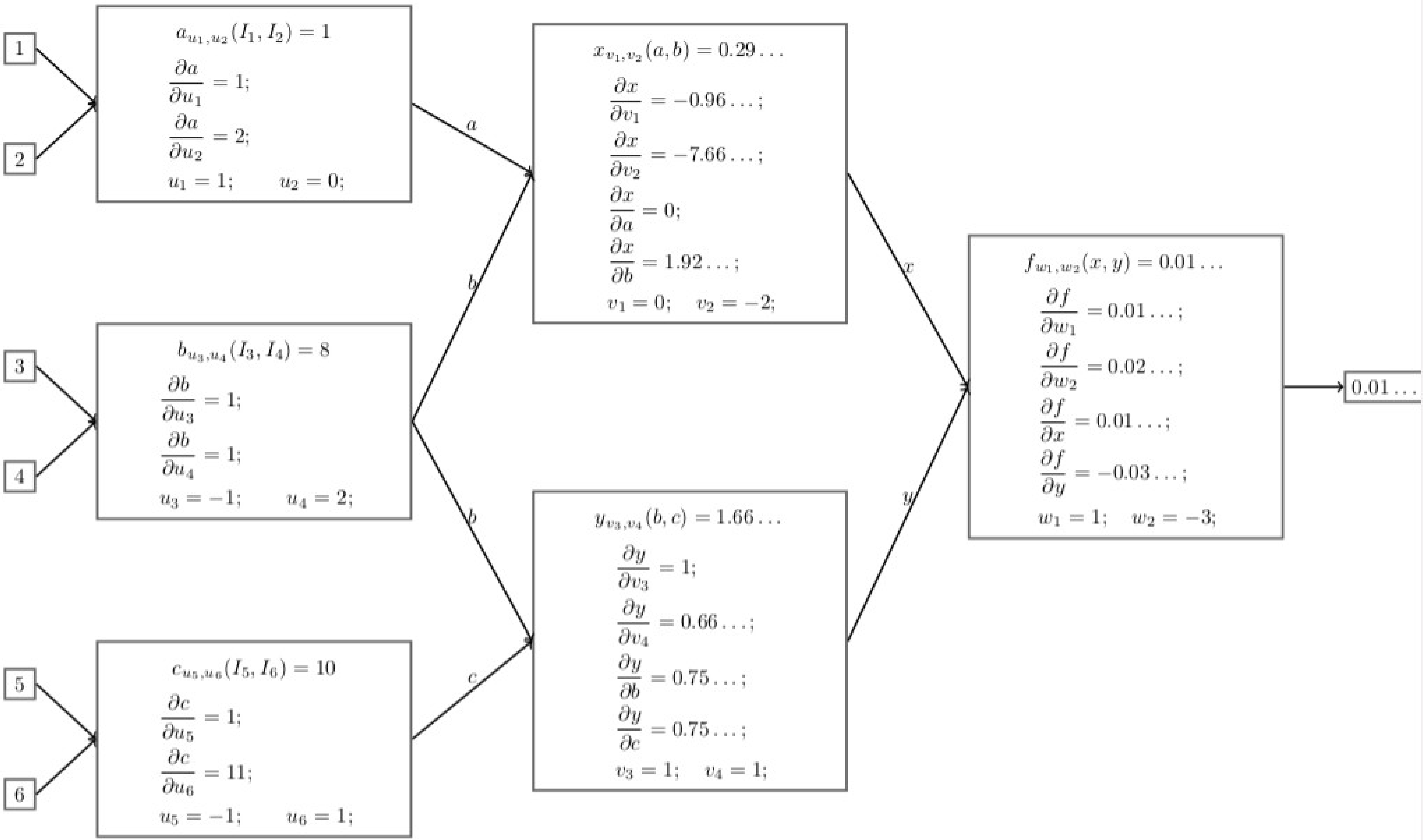
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}}_{\Delta v_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}}_{\Delta v_2}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}}_{\Delta v_3}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}}_{\Delta v_4}, \underbrace{\frac{\partial f}{\partial w_1}}_{\Delta w_1}, \underbrace{\frac{\partial f}{\partial w_2}}_{\Delta w_2} \right\}$$



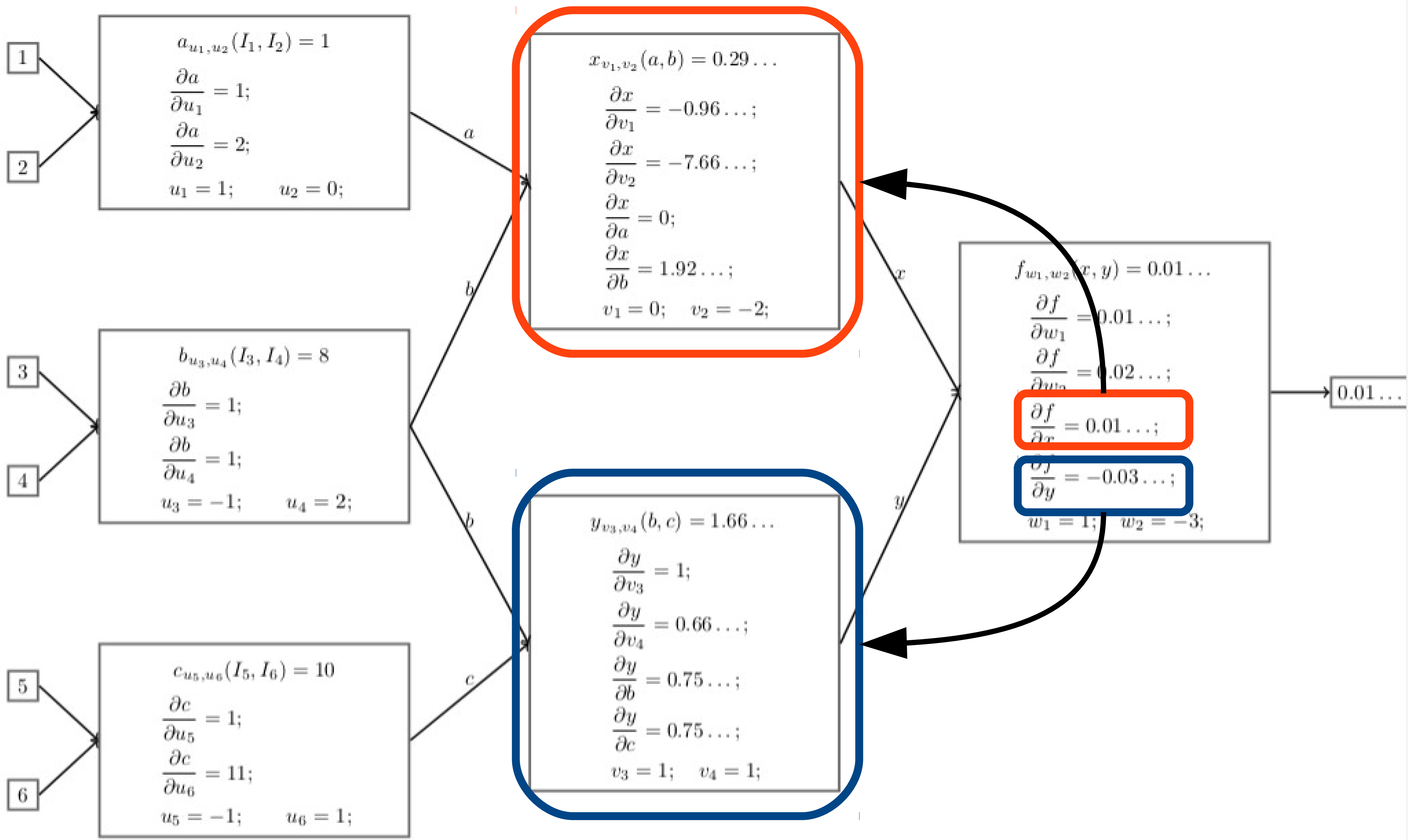
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}}_{\Delta v_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}}_{\Delta v_2}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}}_{\Delta v_3}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$$



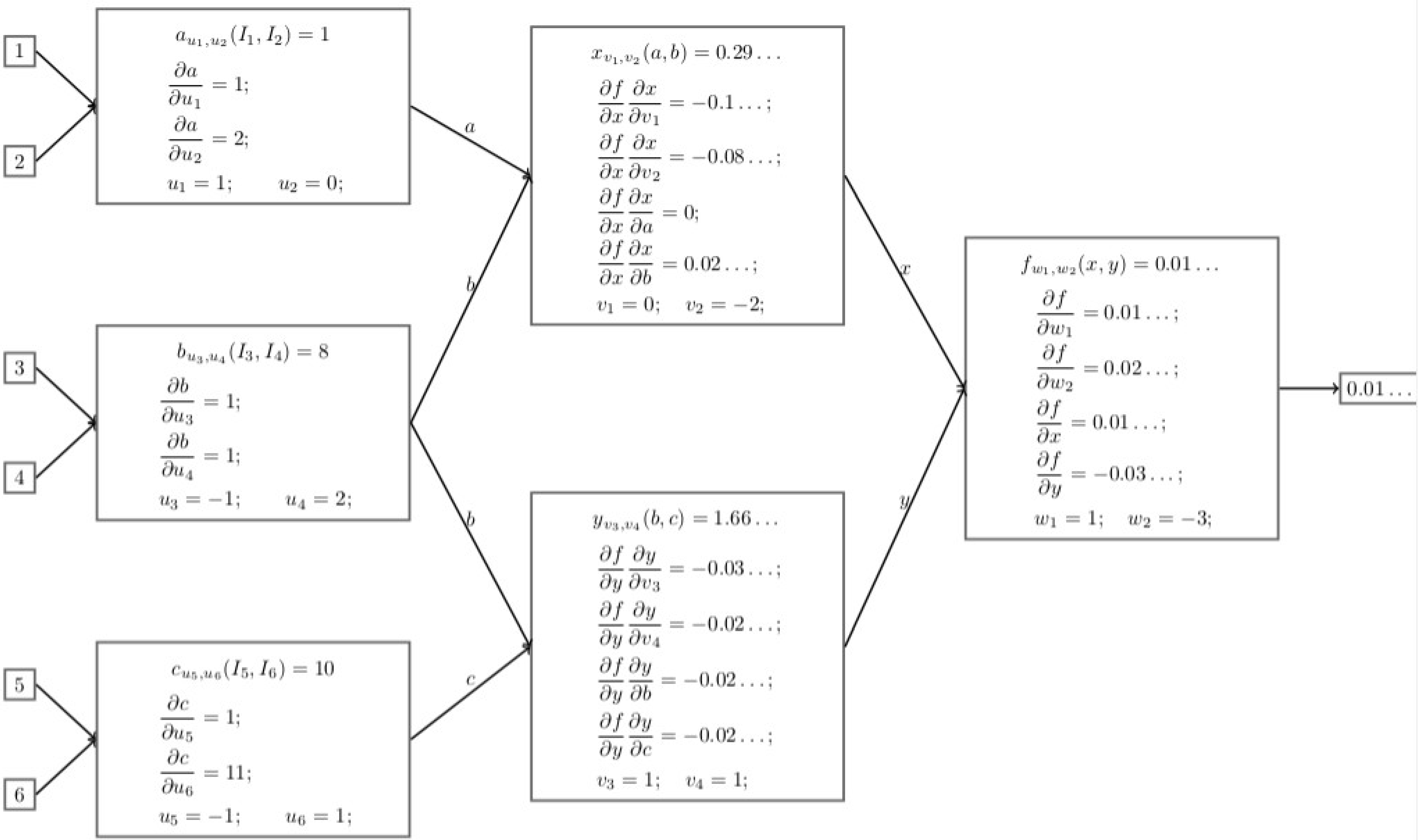
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}}_{\Delta v_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}}_{\Delta v_2}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}}_{\Delta v_3}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$$



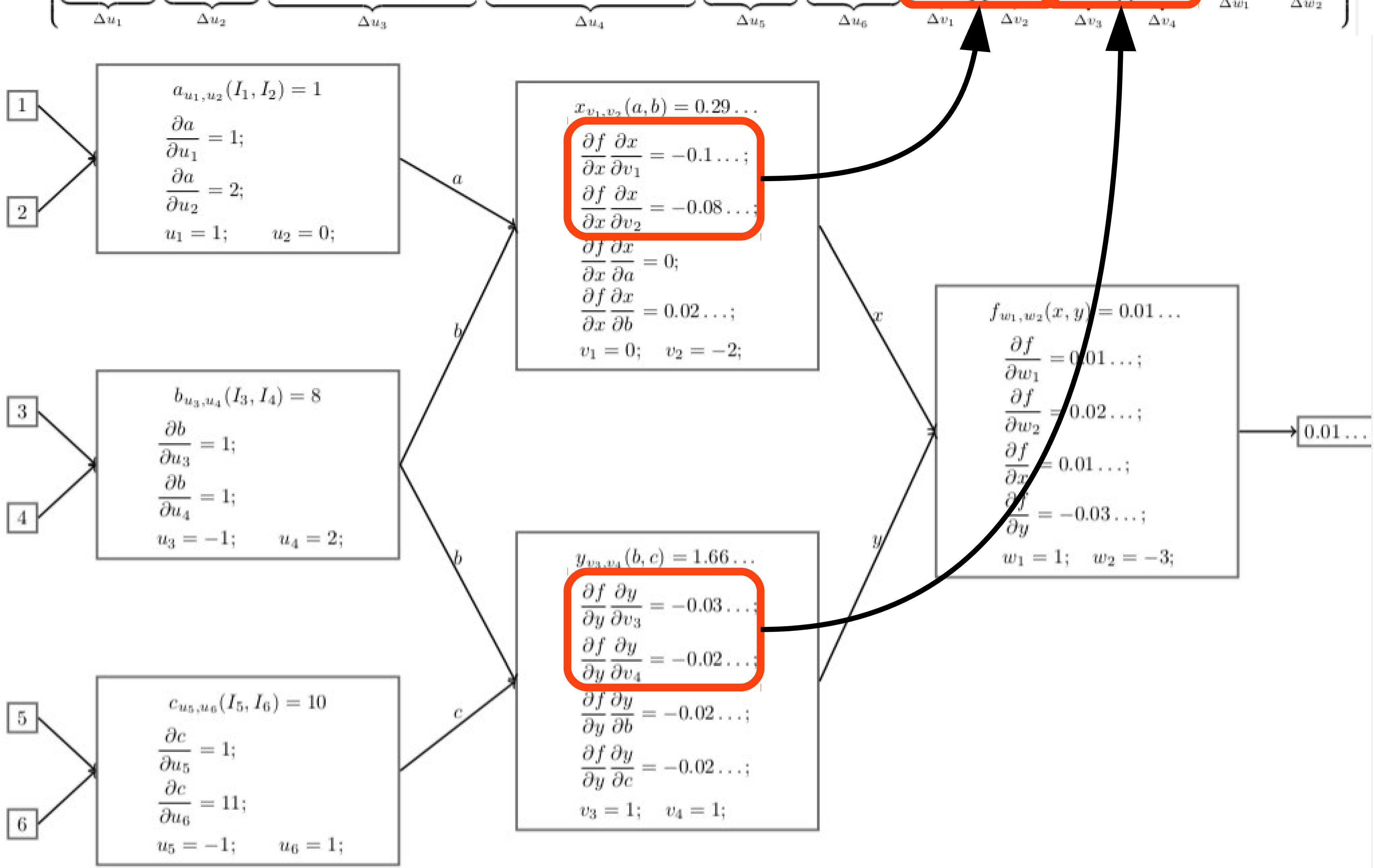
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}}_{\Delta v_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}}_{\Delta v_2}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}}_{\Delta v_3}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$$



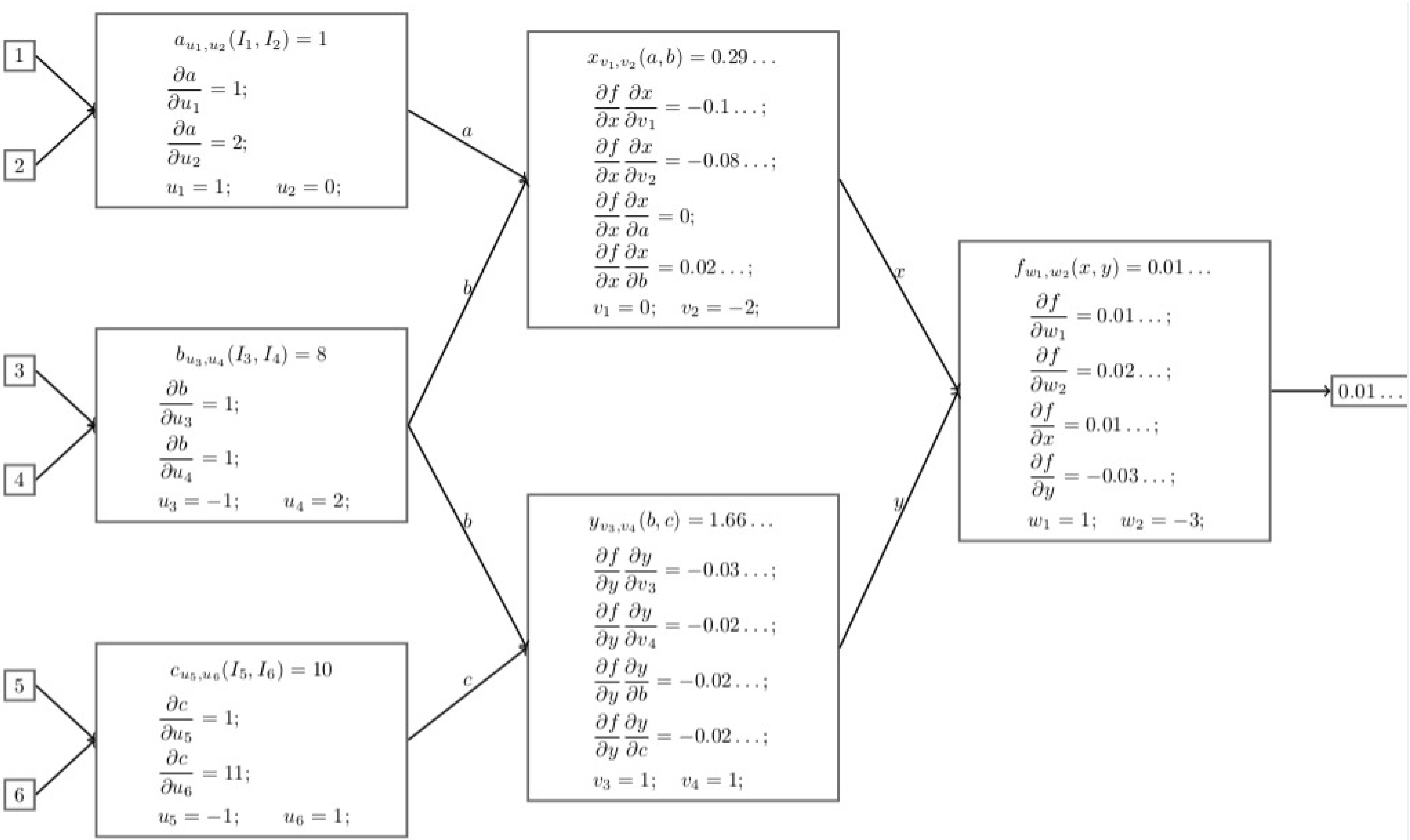
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1}}_{\Delta v_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2}}_{\Delta v_2}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3}}_{\Delta v_3}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4}}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$$



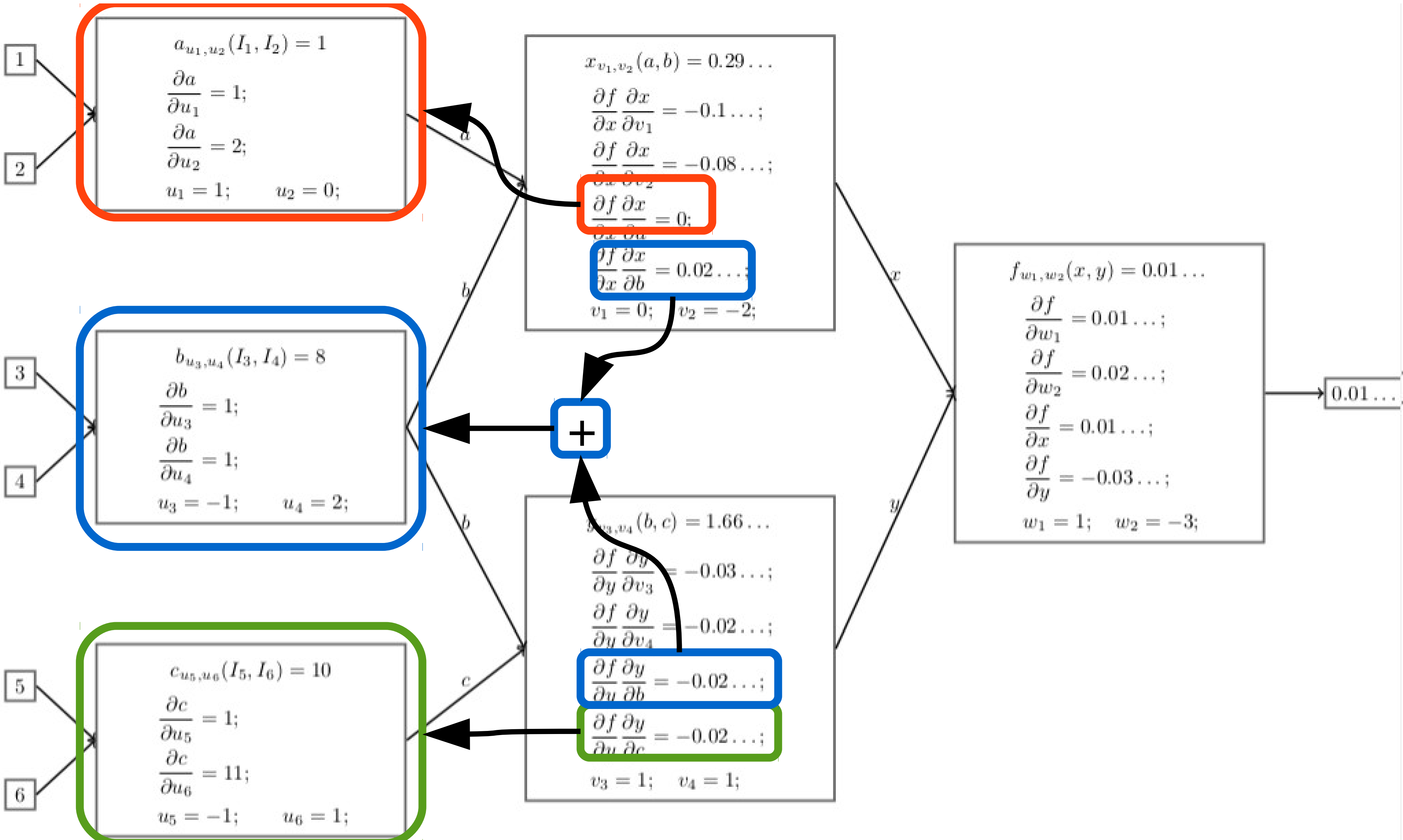
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{-0.1\ldots}_{\Delta v_1}, \underbrace{-0.08\ldots}_{\Delta v_2}, \underbrace{-0.03\ldots}_{\Delta v_3}, \underbrace{-0.02\ldots}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$$



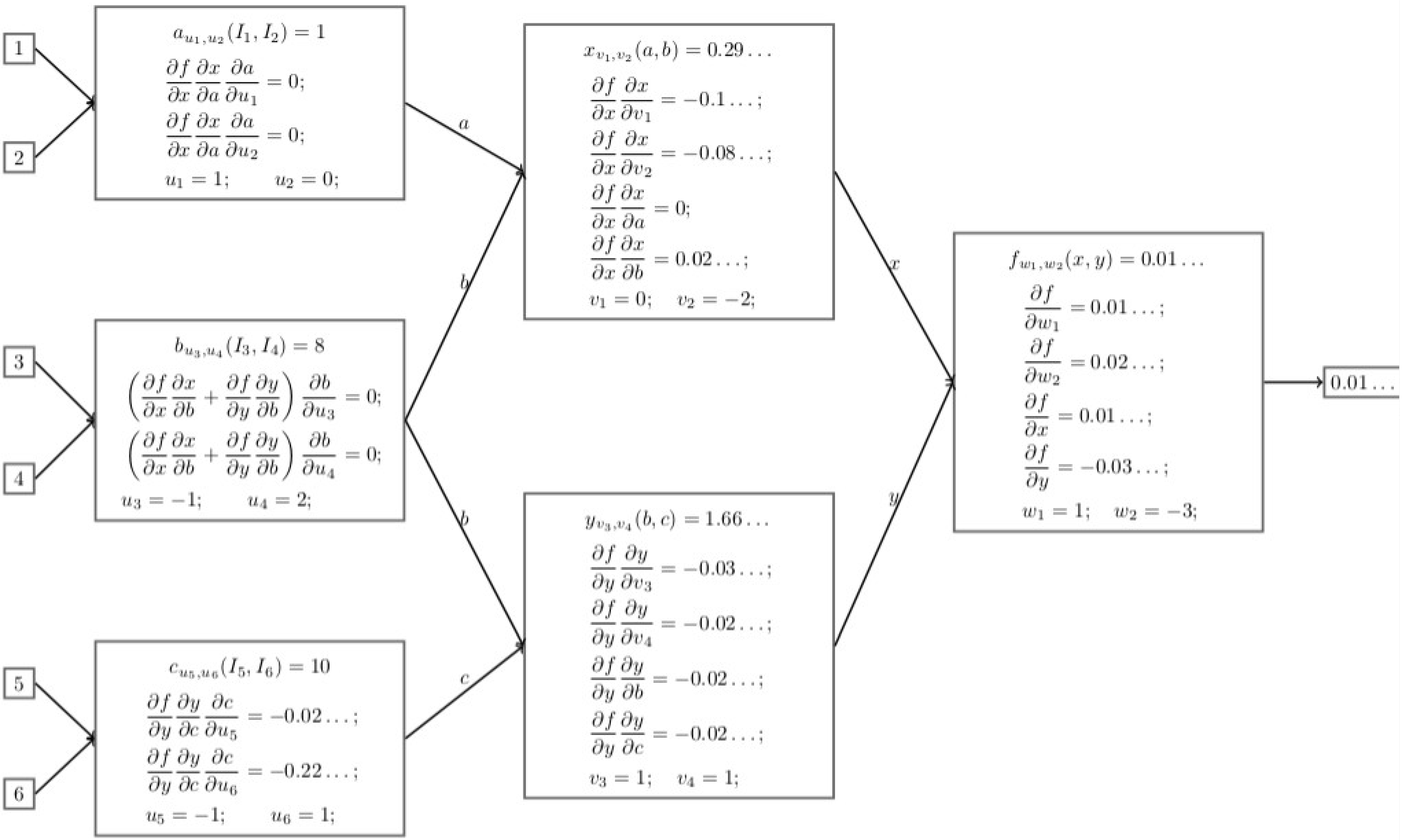
Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{-0.1\ldots}_{\Delta v_1}, \underbrace{-0.08\ldots}_{\Delta v_2}, \underbrace{-0.03\ldots}_{\Delta v_3}, \underbrace{-0.02\ldots}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$$

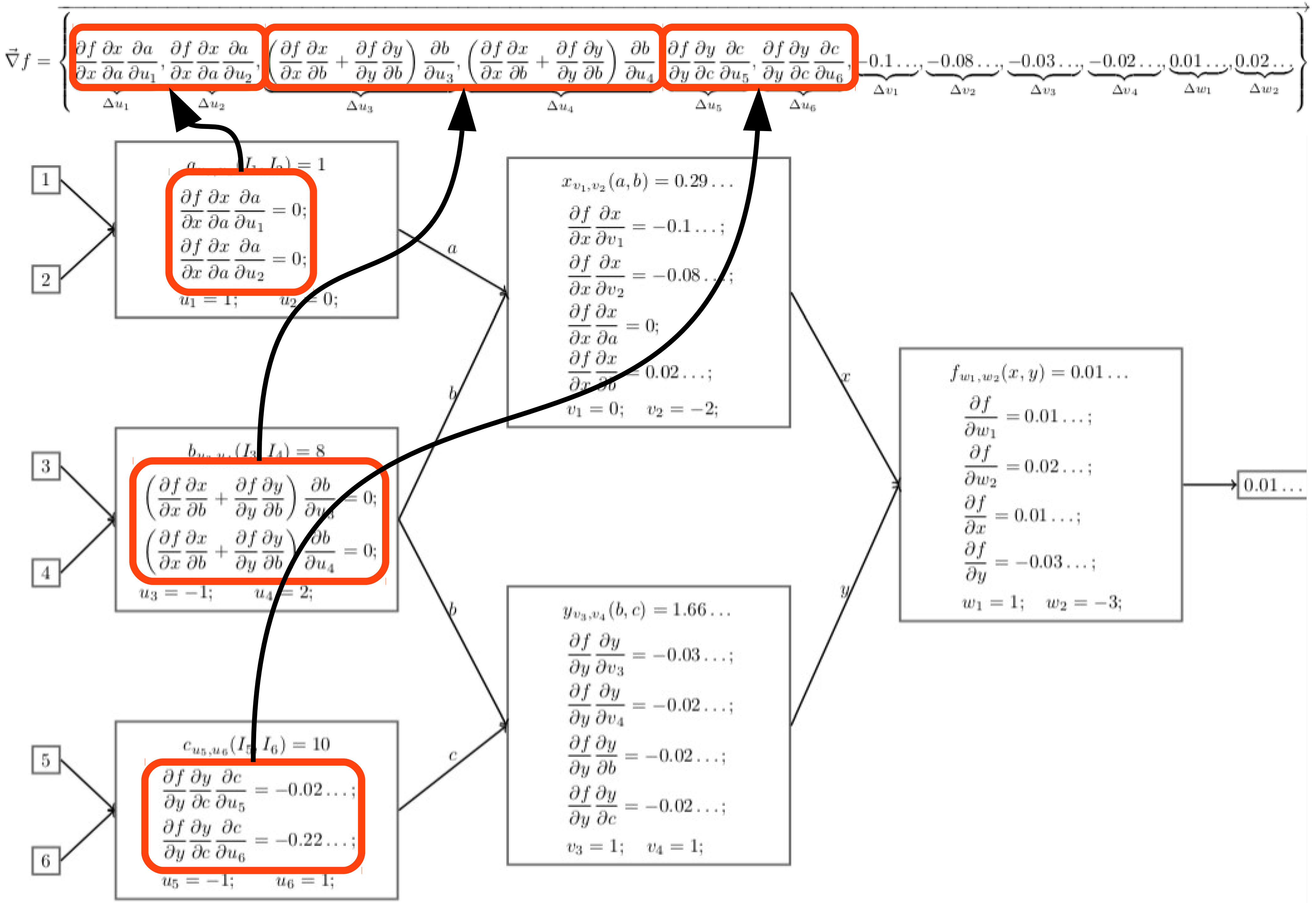


Backward

$$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{-0.1\ldots}_{\Delta v_1}, \underbrace{-0.08\ldots}_{\Delta v_2}, \underbrace{-0.03\ldots}_{\Delta v_3}, \underbrace{-0.02\ldots}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$$



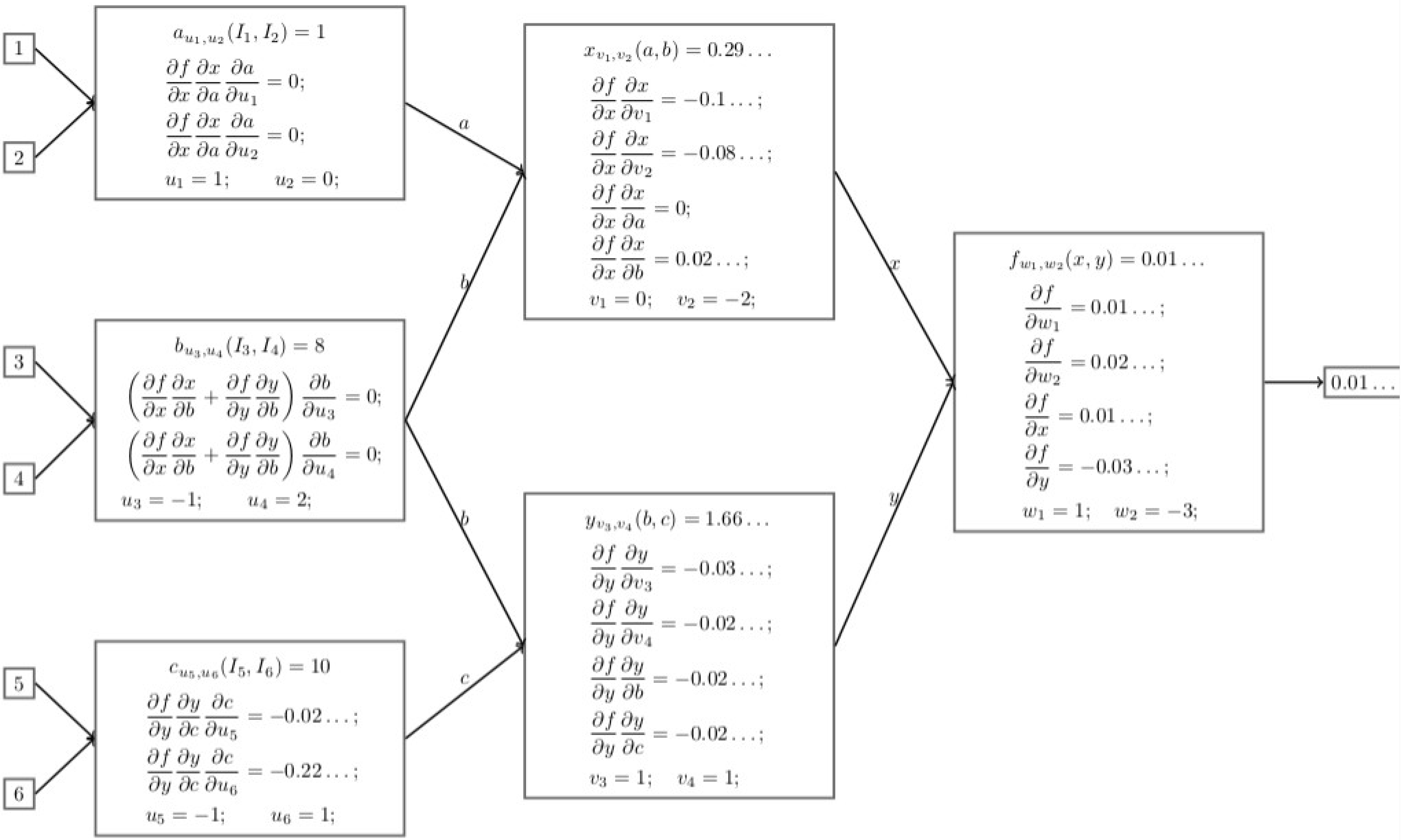
Backward

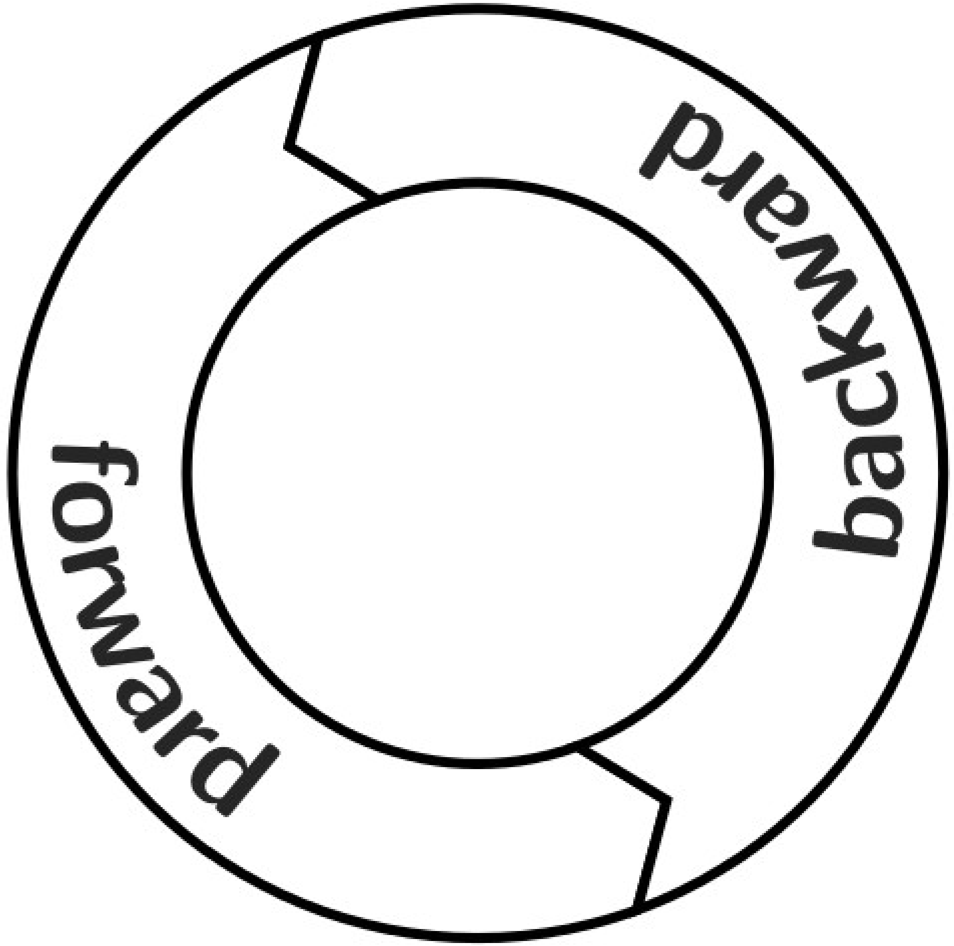


$\vec{\nabla} f = \left\{ \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1}}_{\Delta u_1}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2}}_{\Delta u_2}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3}}_{\Delta u_3}, \underbrace{\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4}}_{\Delta u_4}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5}}_{\Delta u_5}, \underbrace{\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6}}_{\Delta u_6}, \underbrace{-0.1\ldots}_{\Delta v_1}, \underbrace{-0.08\ldots}_{\Delta v_2}, \underbrace{-0.03\ldots}_{\Delta v_3}, \underbrace{-0.02\ldots}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \right\}$
1
2
3
4
5
6
$\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_1} = 0;$
$\frac{\partial f}{\partial x}\frac{\partial x}{\partial a}\frac{\partial a}{\partial u_2} = 0;$
$u_1 = 1;$ $u_2 = 0;$
$b_{u_3,u_4}(I_3, I_4) = 8$
$\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_3} = 0;$
$\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial b}\right)\frac{\partial b}{\partial u_4} = 0;$
$u_3 = -1;$ $u_4 = 2;$
$c_{u_5,u_6}(I_5, I_6) = 10$
$\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_5} = -0.02\ldots;$
$\frac{\partial f}{\partial y}\frac{\partial y}{\partial c}\frac{\partial c}{\partial u_6} = -0.22\ldots;$
$u_5 = -1;$ $u_6 = 1;$
a
b
b
c
$x_{v_1,v_2}(a,b) = 0.29\ldots$
$\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_1} = -0.1\ldots;$
$\frac{\partial f}{\partial x}\frac{\partial x}{\partial v_2} = -0.08\ldots;$
$\frac{\partial f}{\partial x}\frac{\partial x}{\partial a} = 0;$
$\frac{\partial f}{\partial x}\frac{\partial x}{\partial b} = 0.02\ldots;$
$v_1 = 0;$ $v_2 = -2;$
$y_{v_3,v_4}(b,c) = 1.66\ldots$
$\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_3} = -0.03\ldots;$
$\frac{\partial f}{\partial y}\frac{\partial y}{\partial v_4} = -0.02\ldots;$
$\frac{\partial f}{\partial y}\frac{\partial y}{\partial b} = -0.02\ldots;$
$\frac{\partial f}{\partial y}\frac{\partial y}{\partial c} = -0.02\ldots;$
$v_3 = 1;$ $v_4 = 1;$
x
y
$f_{w_1,w_2}(x,y) = 0.01\ldots$
$\frac{\partial f}{\partial w_1} = 0.01\ldots;$
$\frac{\partial f}{\partial w_2} = 0.02\ldots;$
$\frac{\partial f}{\partial x} = 0.01\ldots;$
$\frac{\partial f}{\partial y} = -0.03\ldots;$
$w_1 = 1;$ $w_2 = -3;$
0.01...

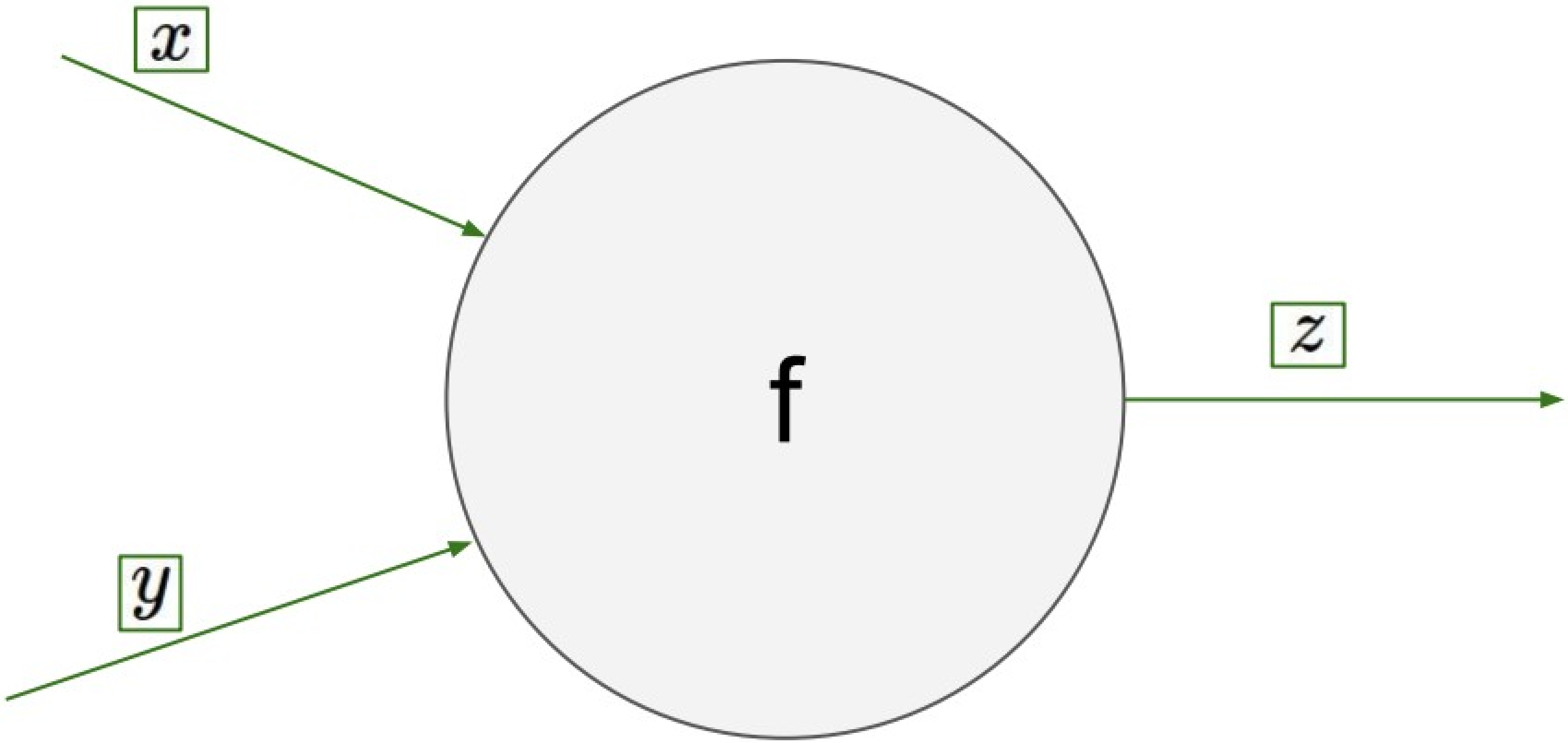
Backward

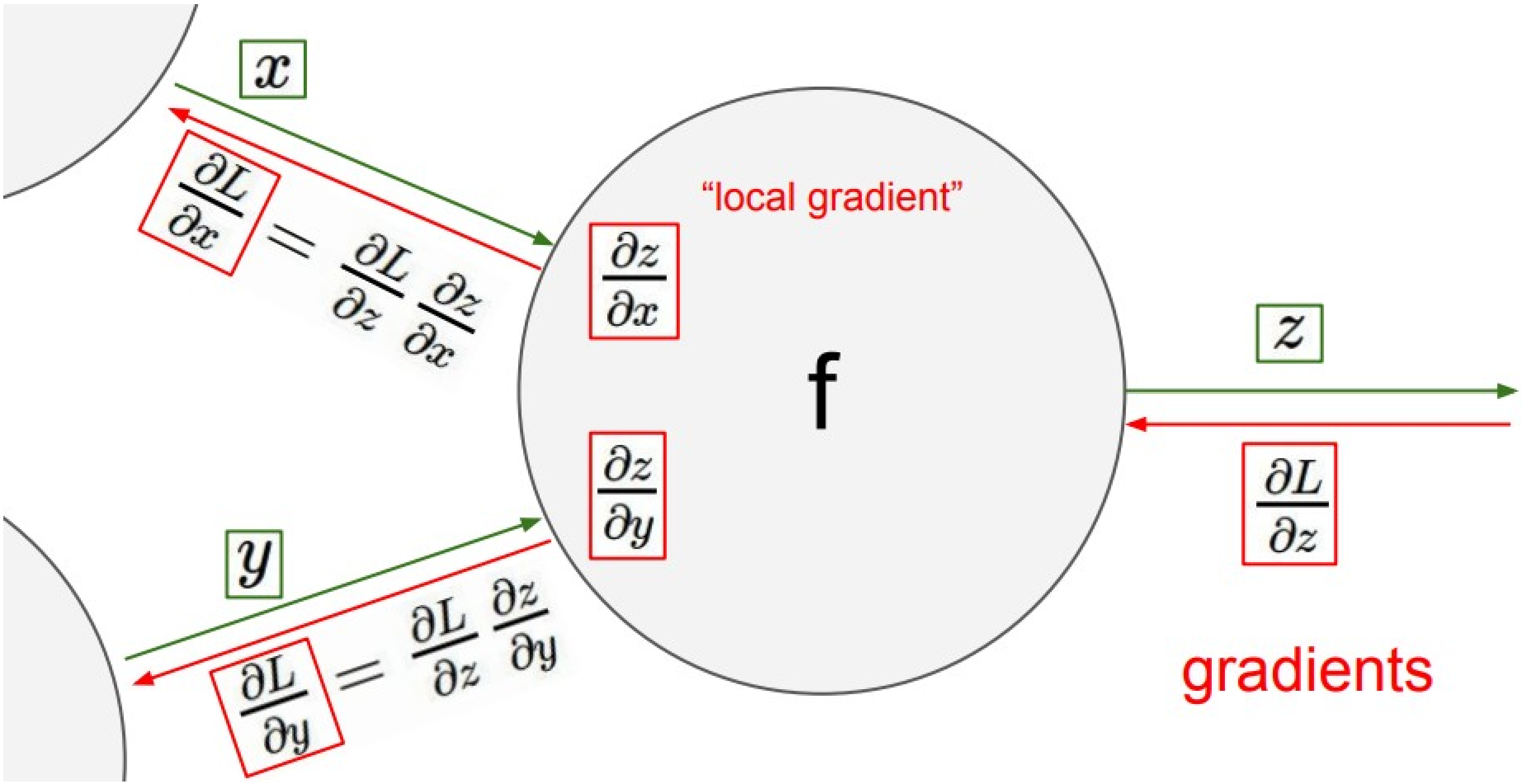
$$\vec{\nabla} f = \Big\{ \underbrace{0}_{\Delta u_1}, \underbrace{0}_{\Delta u_2}, \underbrace{0}_{\Delta u_3}, \underbrace{0}_{\Delta u_4}, \underbrace{-0.02\ldots}_{\Delta u_5}, \underbrace{-0.22\ldots}_{\Delta u_6}, \underbrace{-0.1\ldots}_{\Delta v_1}, \underbrace{-0.08\ldots}_{\Delta v_2}, \underbrace{-0.03\ldots}_{\Delta v_3}, \underbrace{-0.02\ldots}_{\Delta v_4}, \underbrace{0.01\ldots}_{\Delta w_1}, \underbrace{0.02\ldots}_{\Delta w_2} \Big\}$$



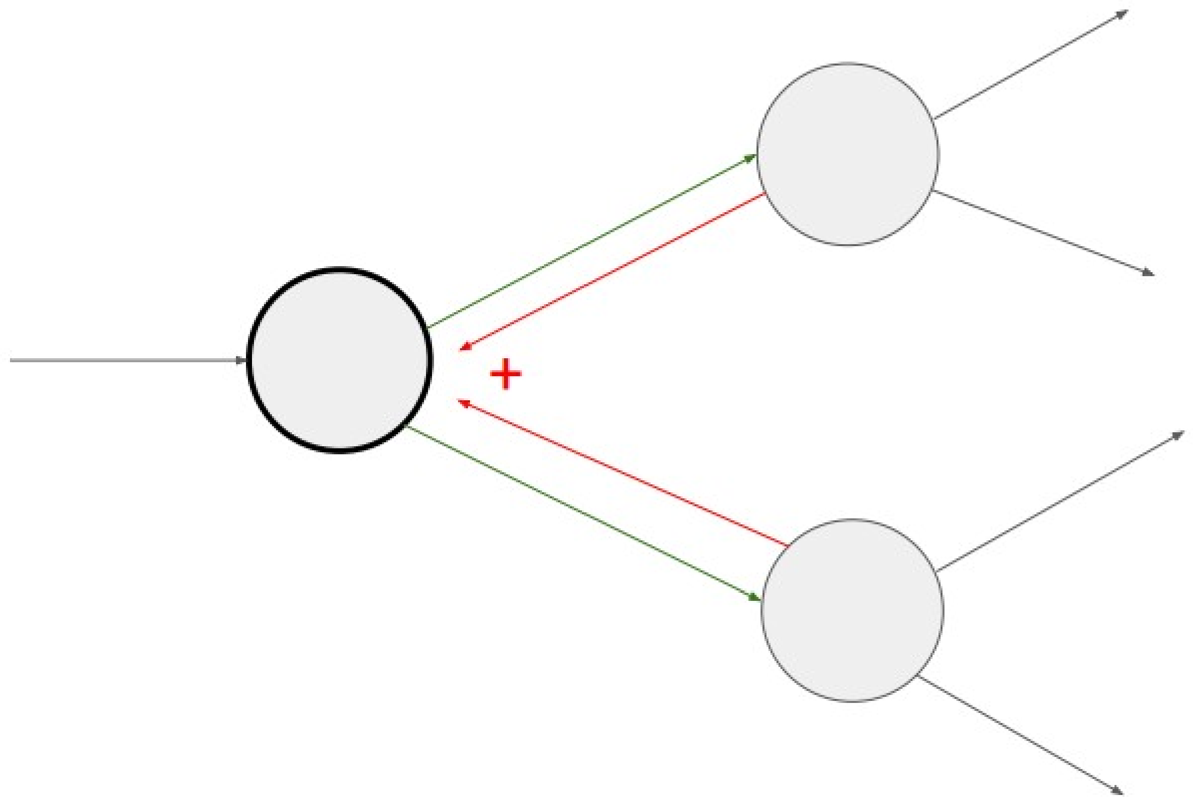


forward
backward


x
y
f
z

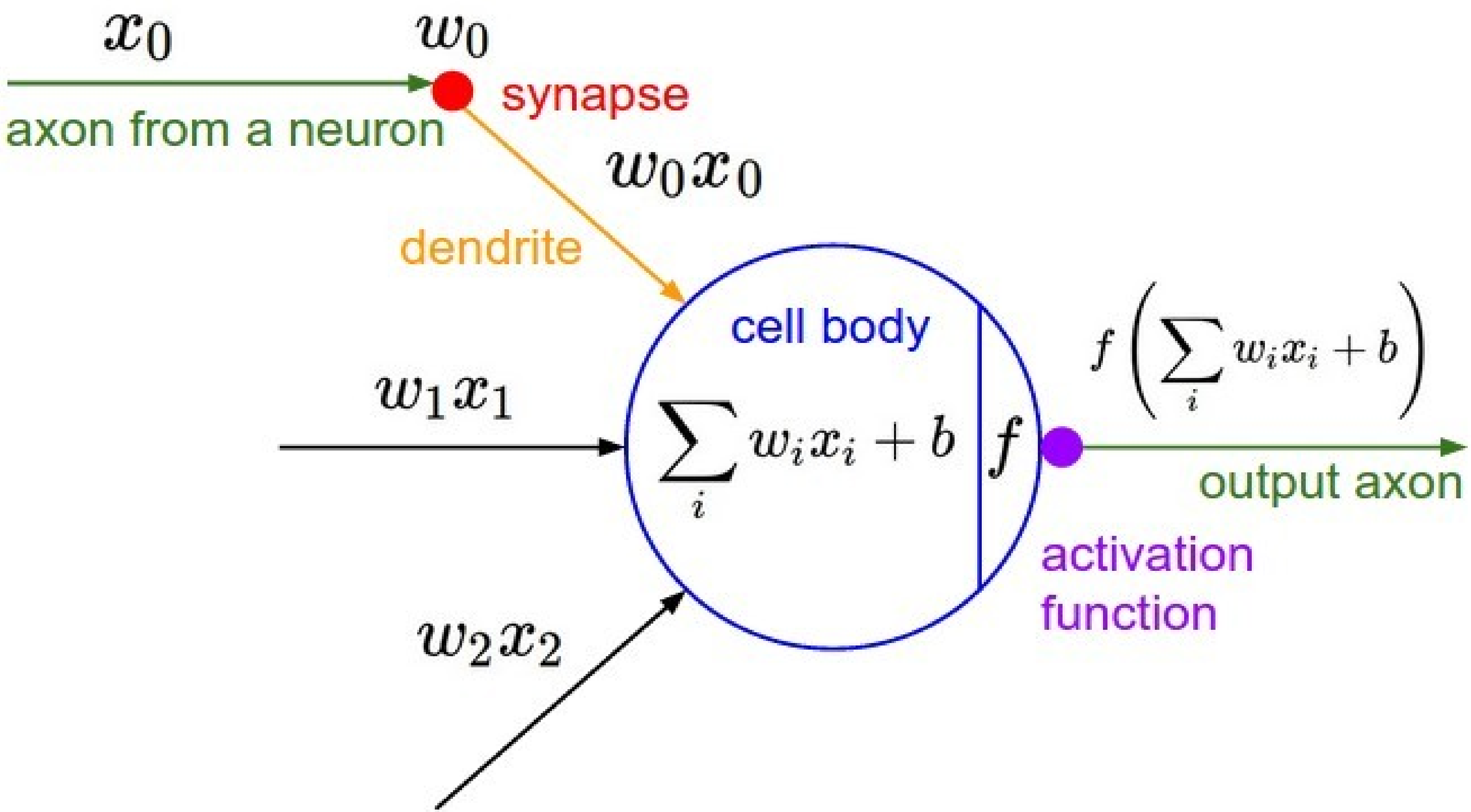

x
$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial x}$
"local gradient"
$\frac{\partial z}{\partial x}$
f
$\frac{\partial z}{\partial y}$
z
$\frac{\partial L}{\partial z}$
y
$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial y}$
gradients

Gradients add at branches



Possible issues with deep networks

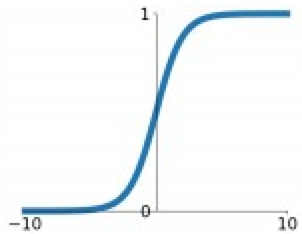
Single Neuron



Activation functions

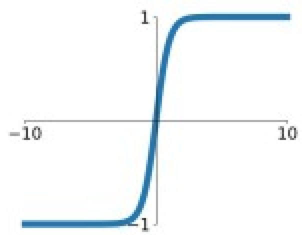
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



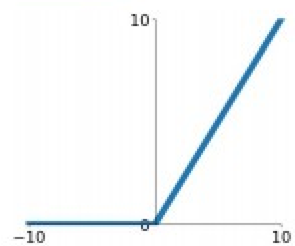
tanh

$\tanh(x)$



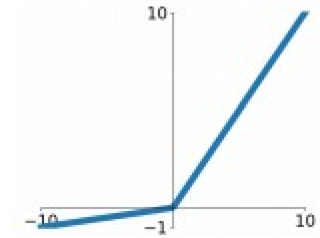
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

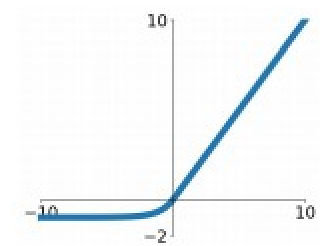


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

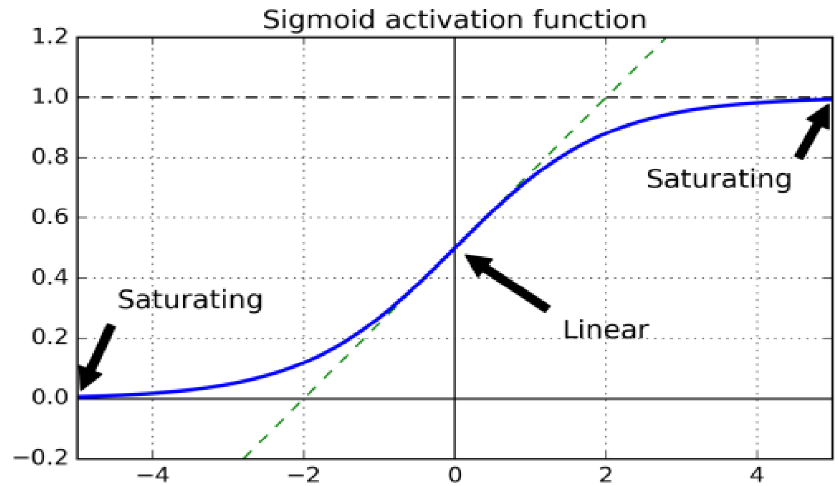
ELU

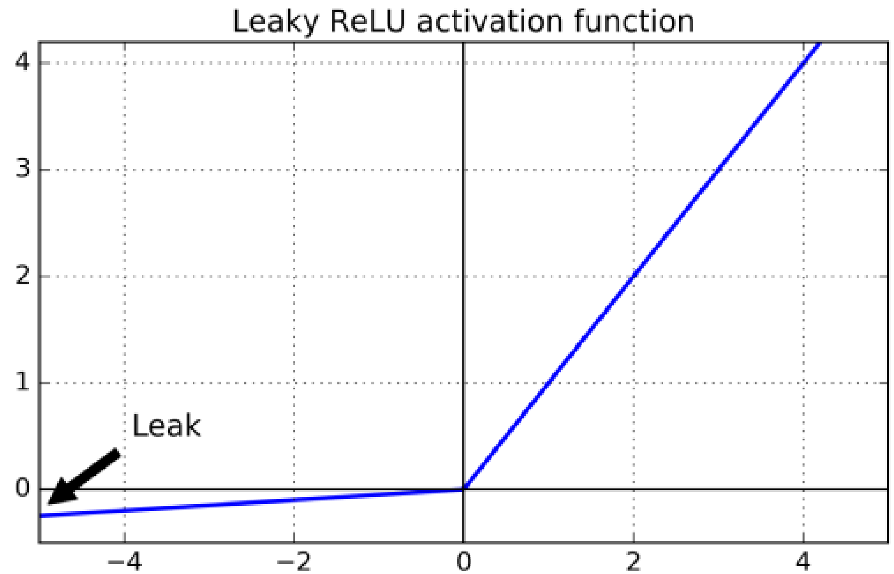
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

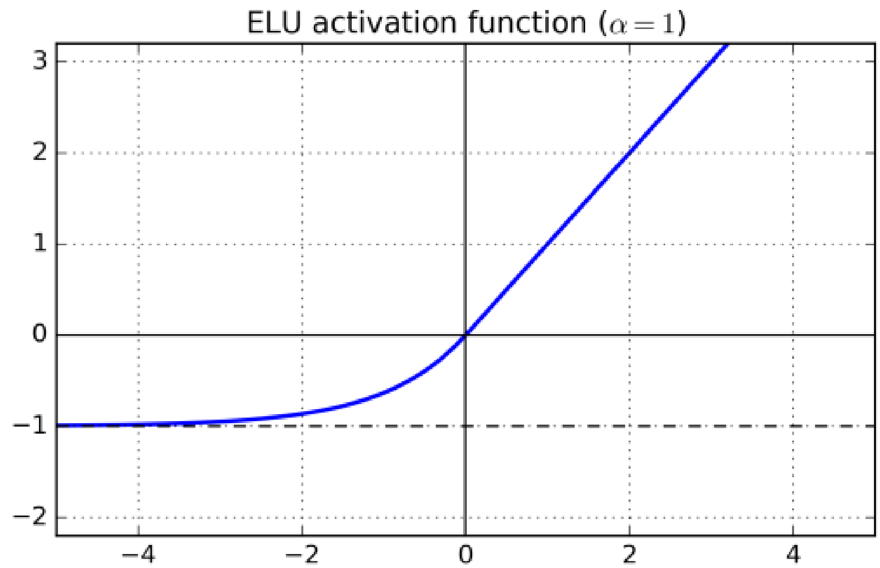


Vanishing/exploding gradients problems

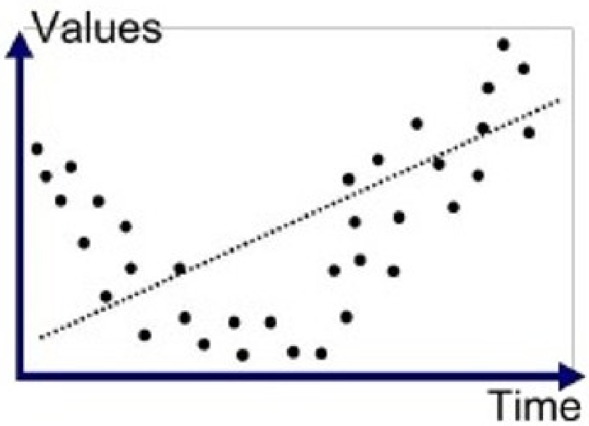
- Xavier and He initialization
- Non-saturating activation functions
- Batch Normalization
- Gradient clipping



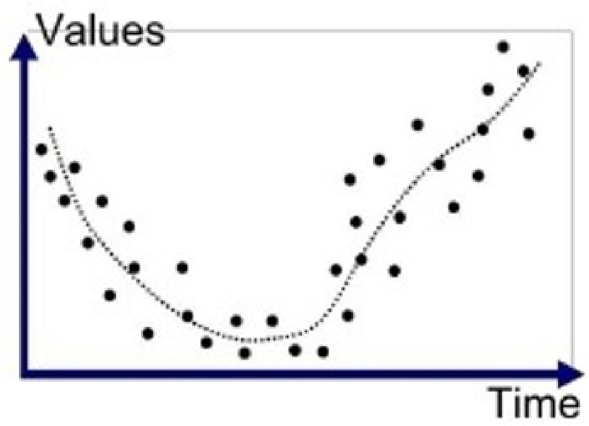




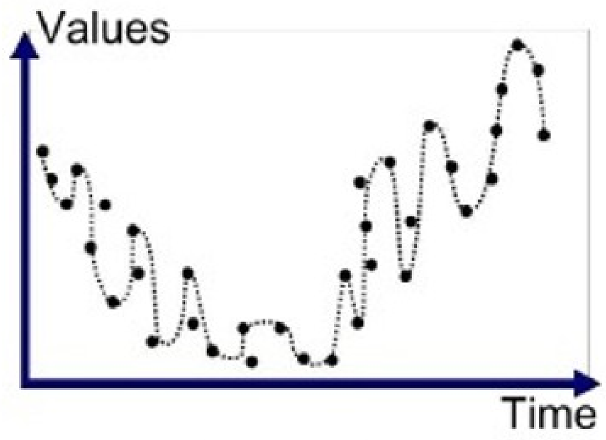
Fitting problems



Underfitted



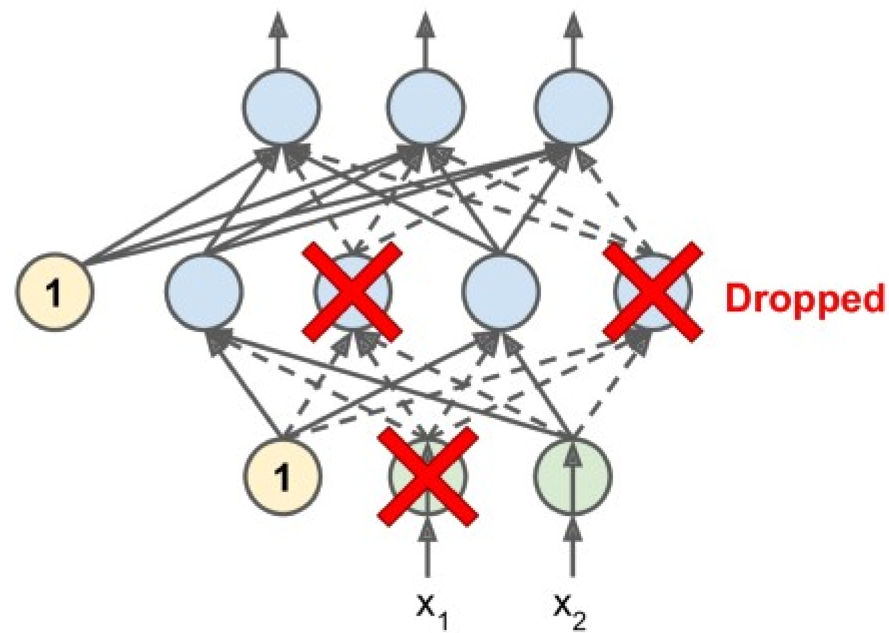
Good Fit/Robust

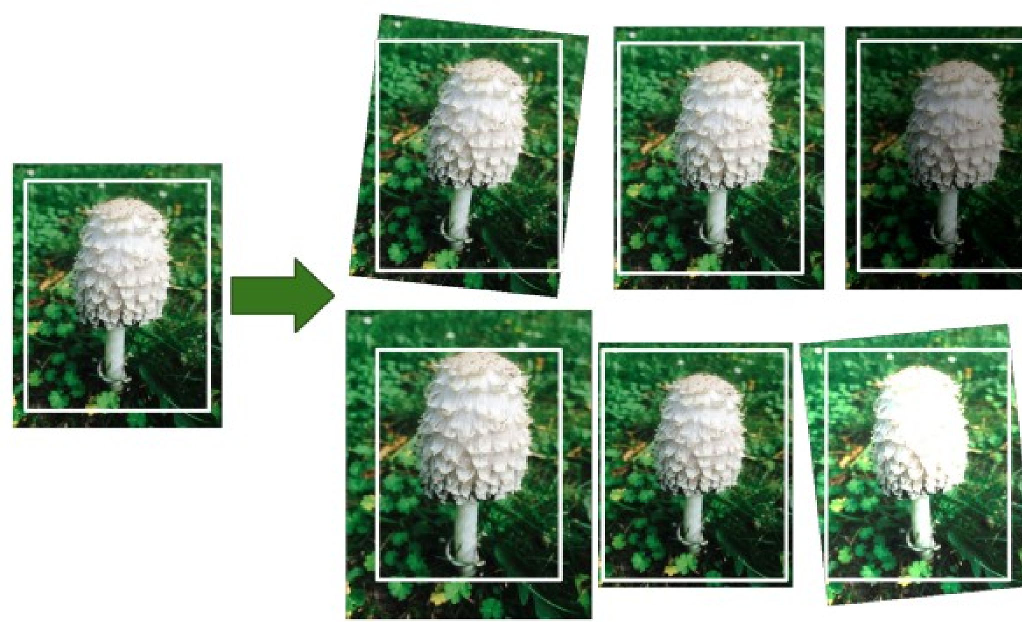


Overfitted

Avoid overfitting

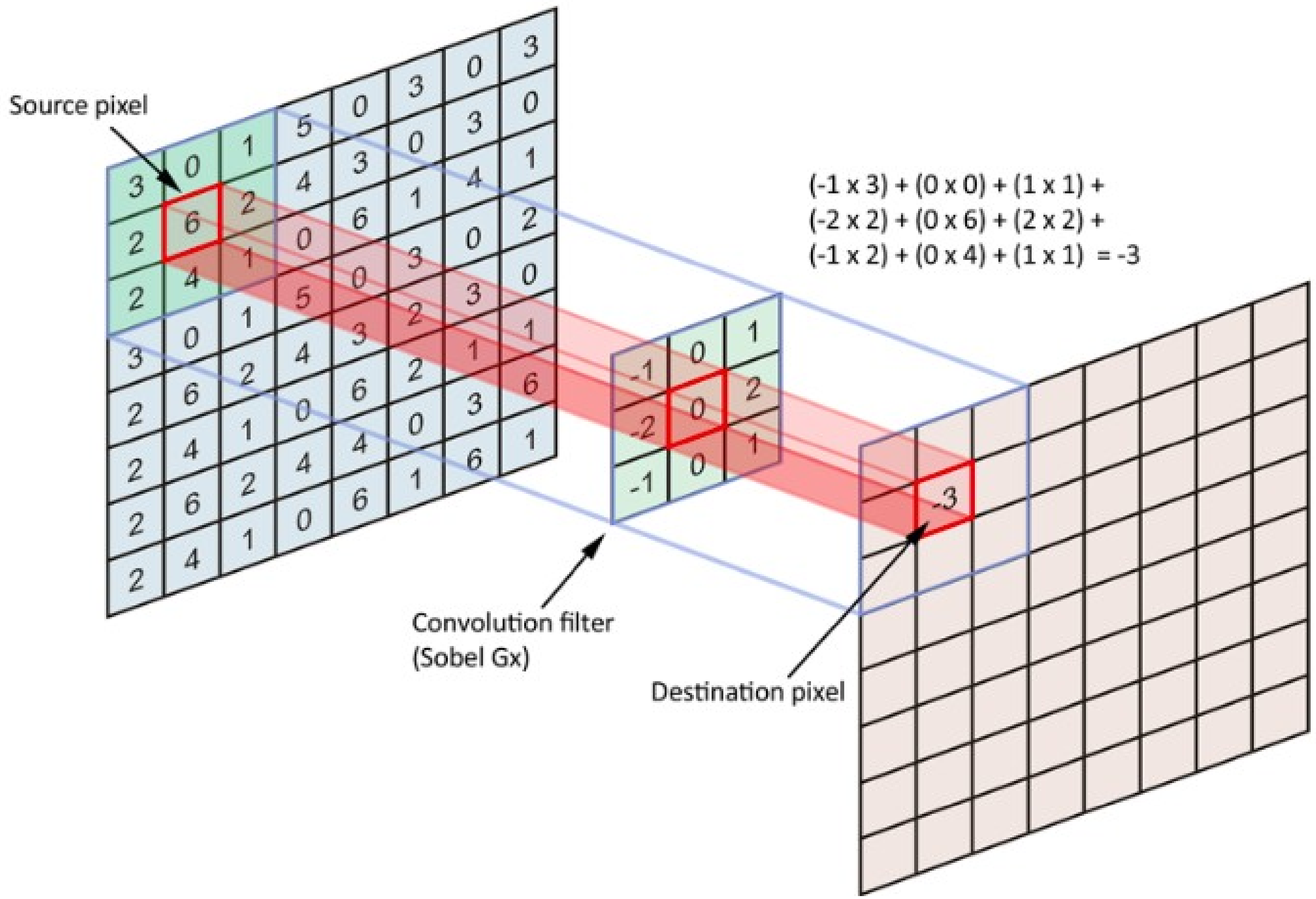
- Early Stopping
- L1 and L2 regularization
- Dropout
- Max-norm regularization
- Data augmentation



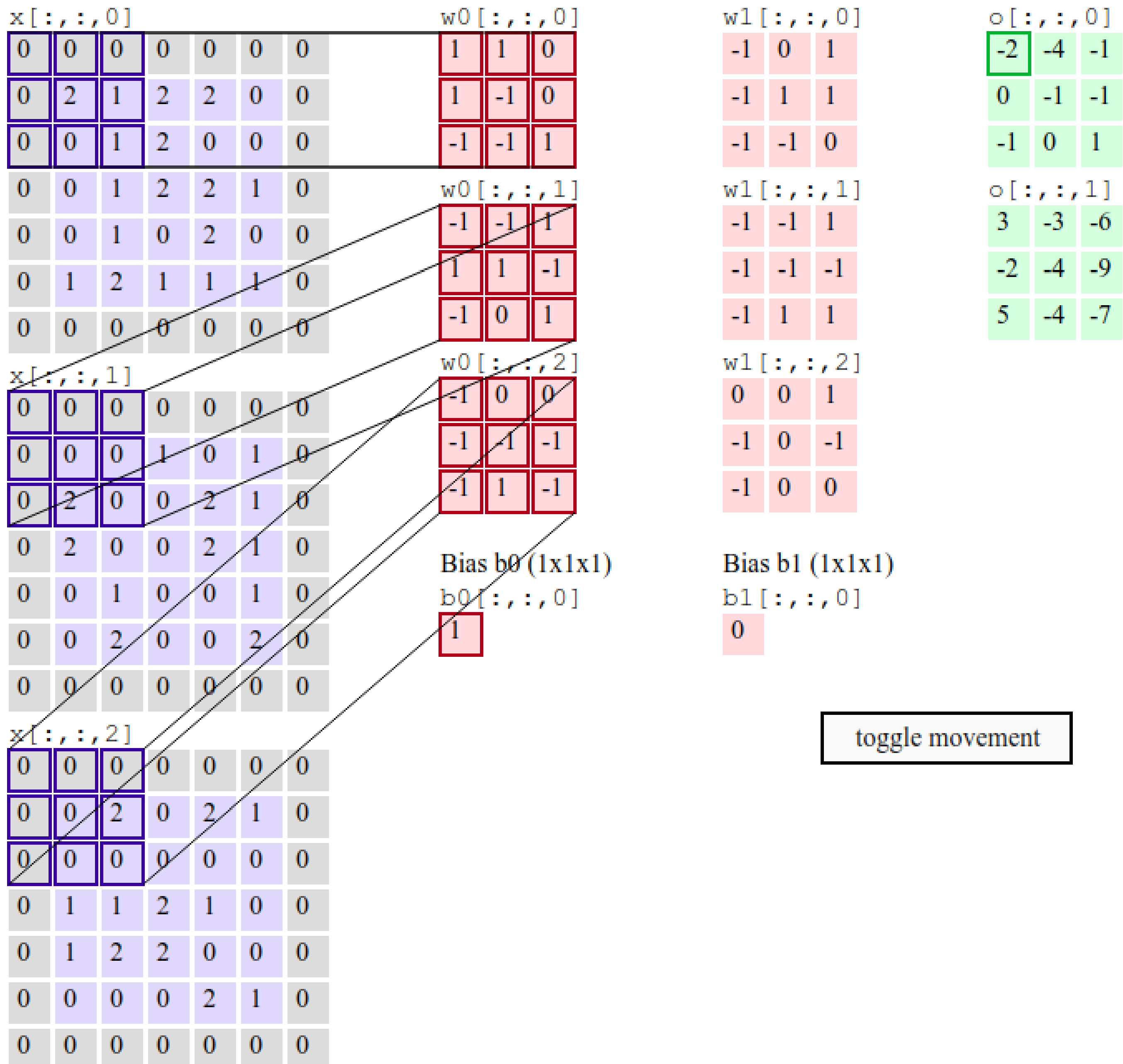


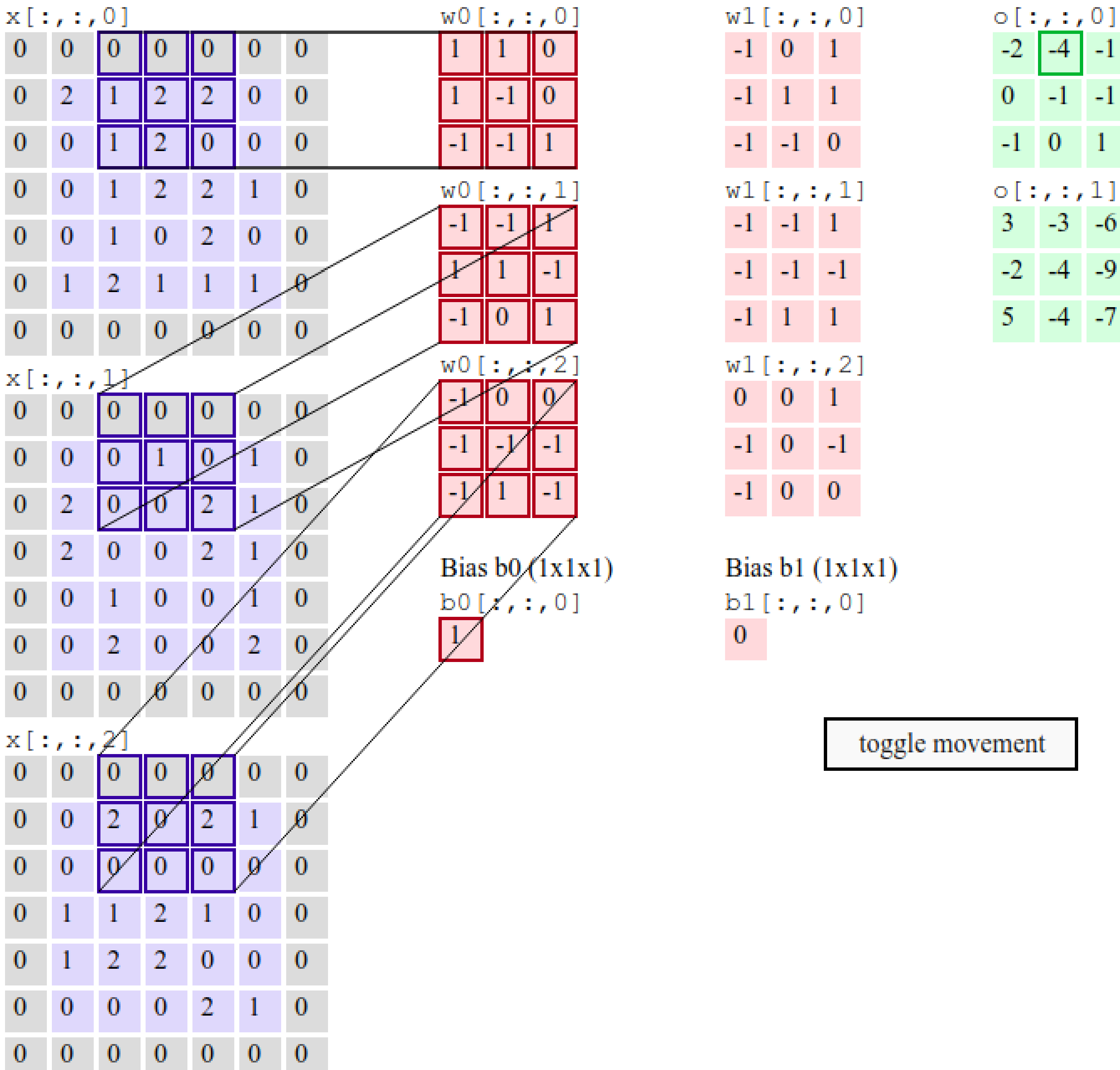
Convolution operation

Convolution

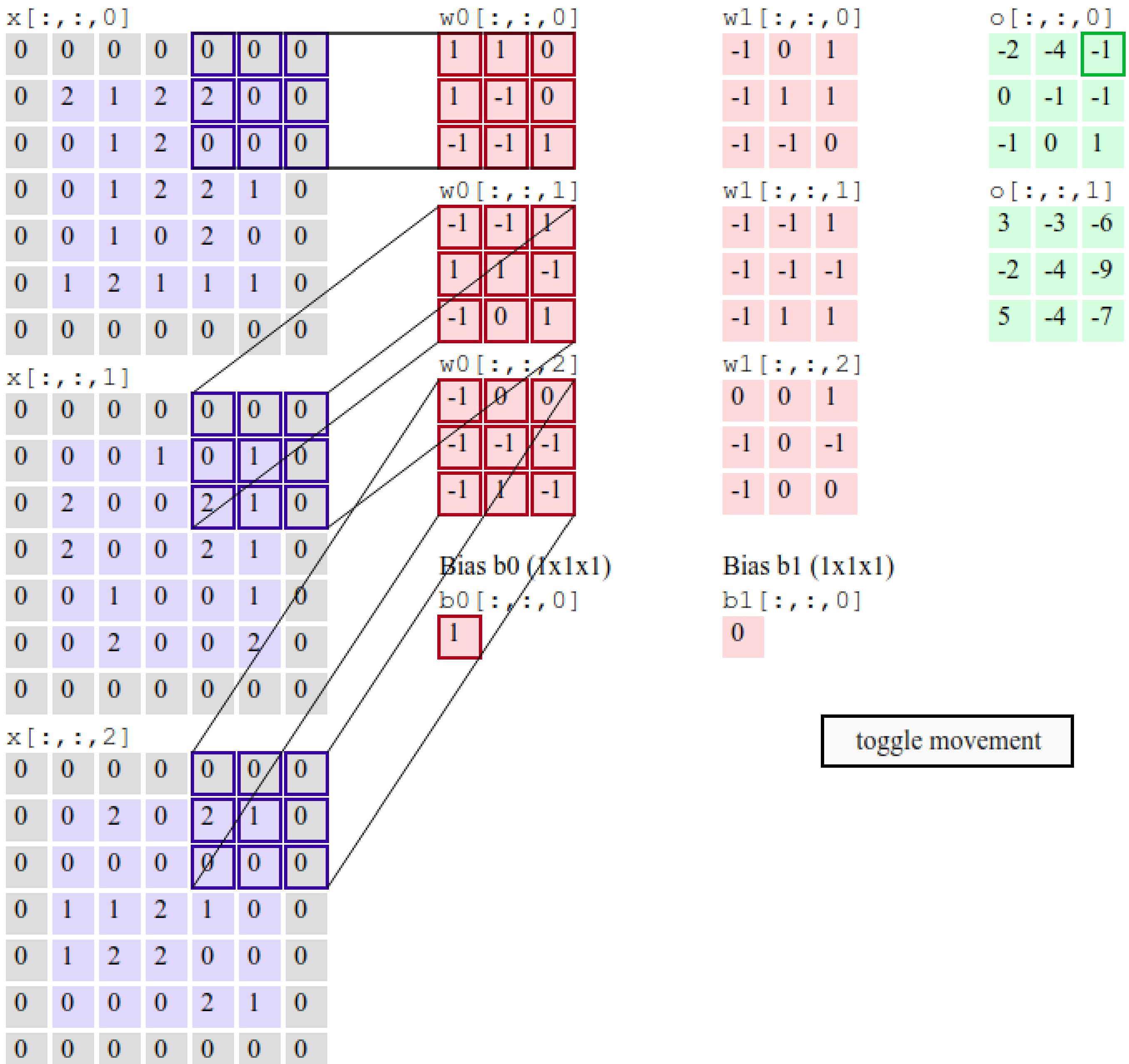


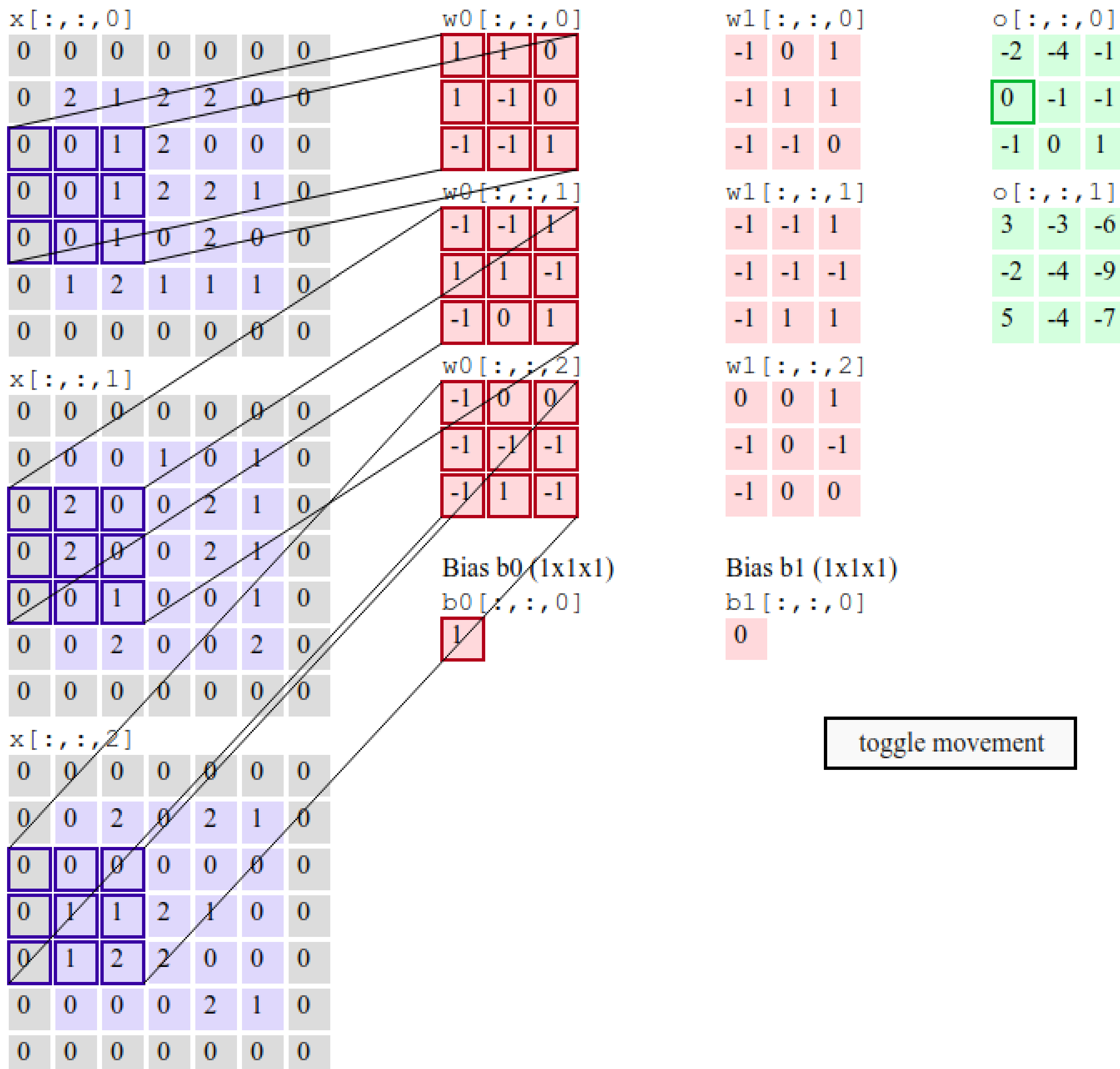
Two basic ideas:
- geometrical proximity has significant meaning
- translational invariance

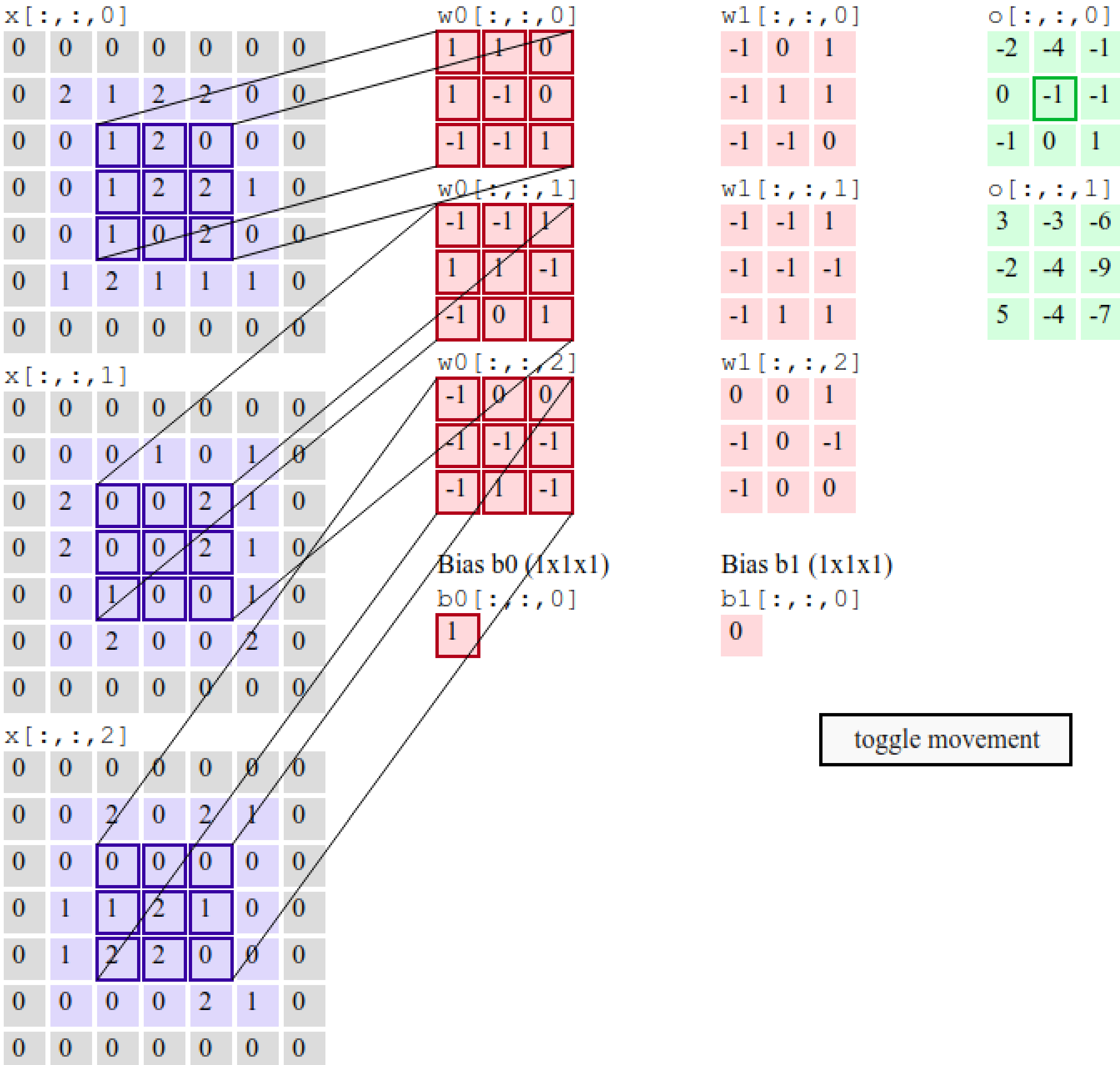

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

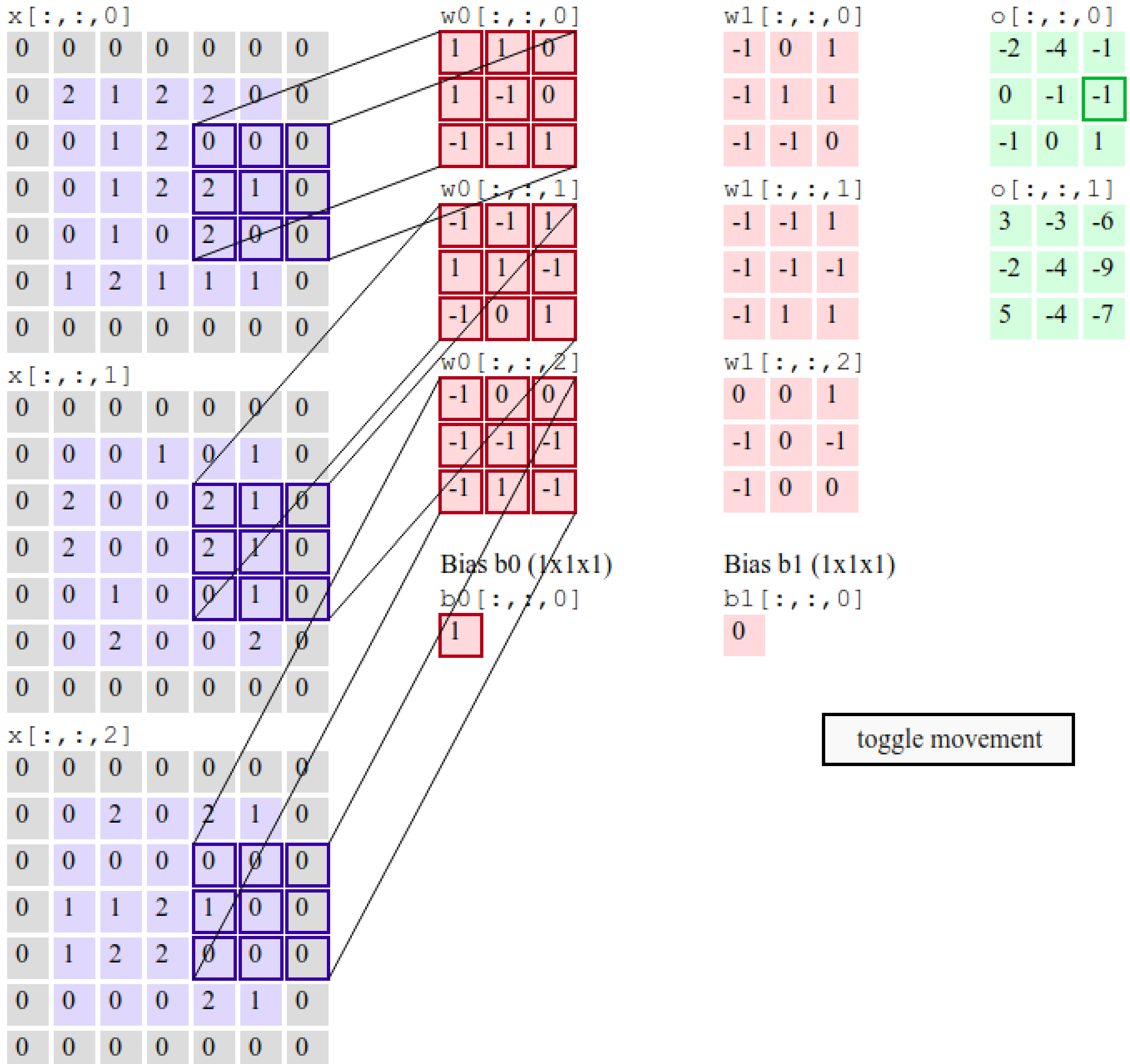


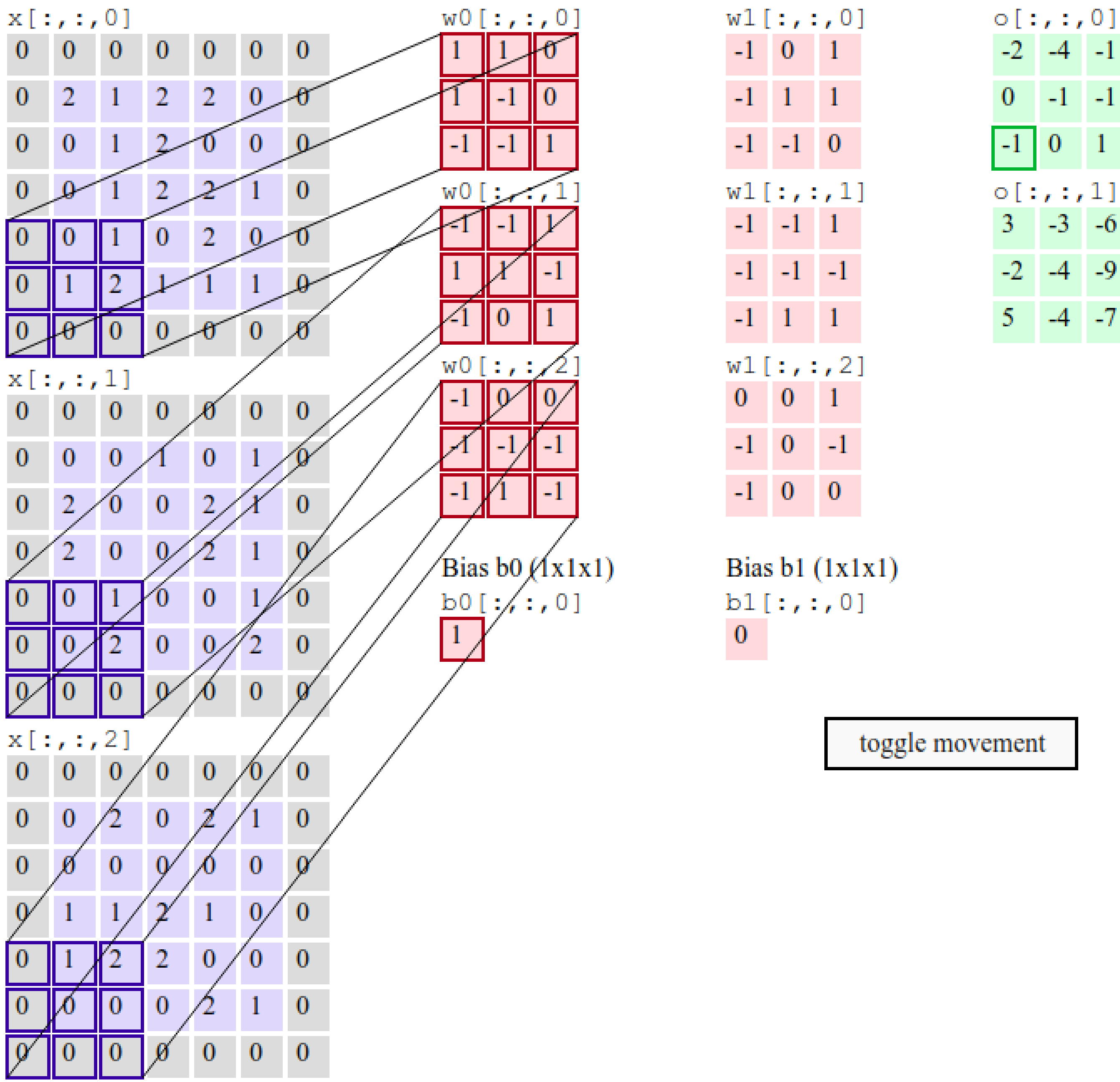
x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

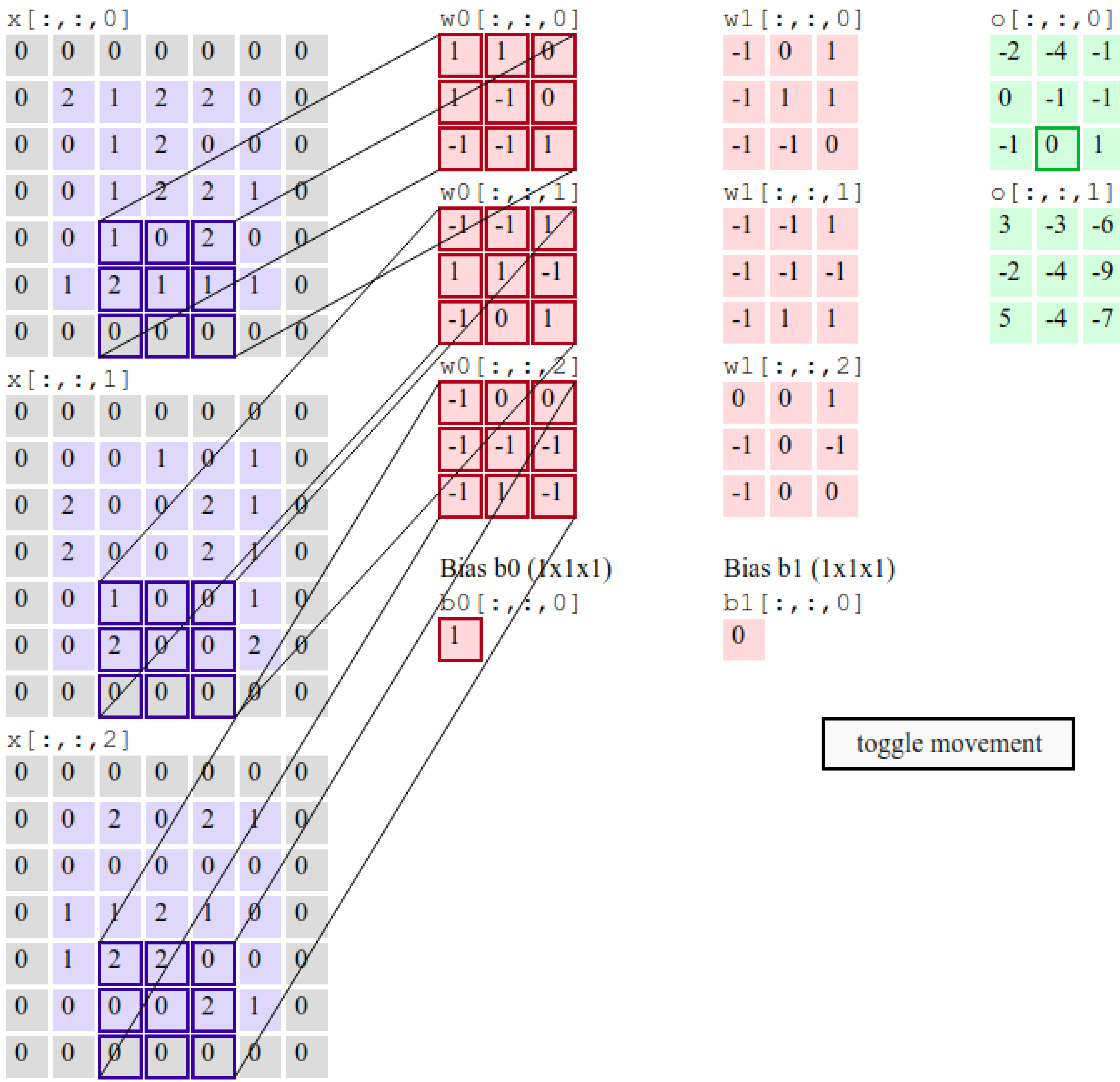

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

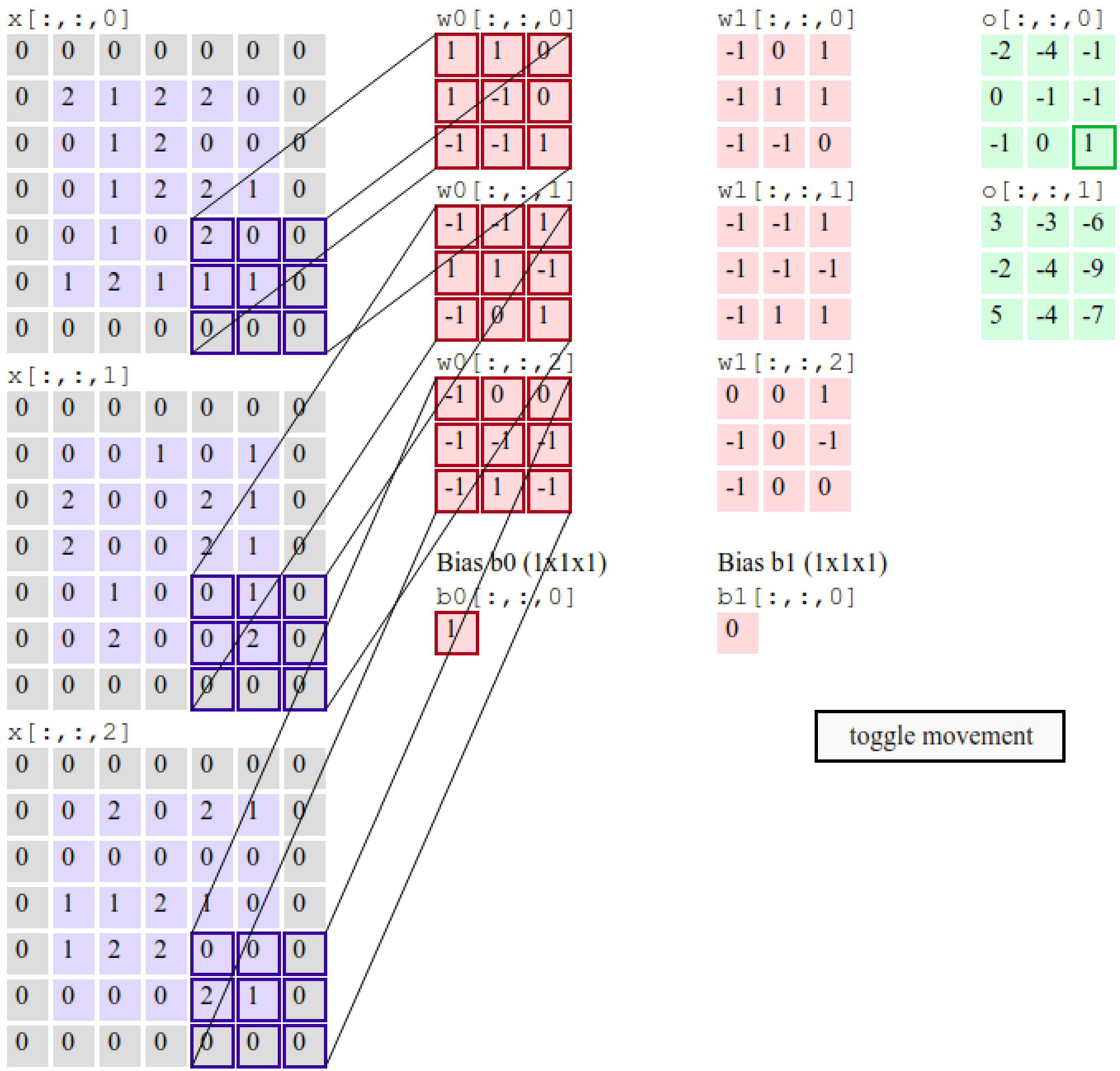

x[:,:,0]
x[:,:,1]
x[:,:,2]
w0[:,:,0]
w0[:,:,1]
w0[:,:,2]
Bias b0 (1x1x1)
b0[:,:,0]
w1[:,:,0]
w1[:,:,1]
w1[:,:,2]
Bias b1 (1x1x1)
b1[:,:,0]
o[:,:,0]
o[:,:,1]
toggle movement

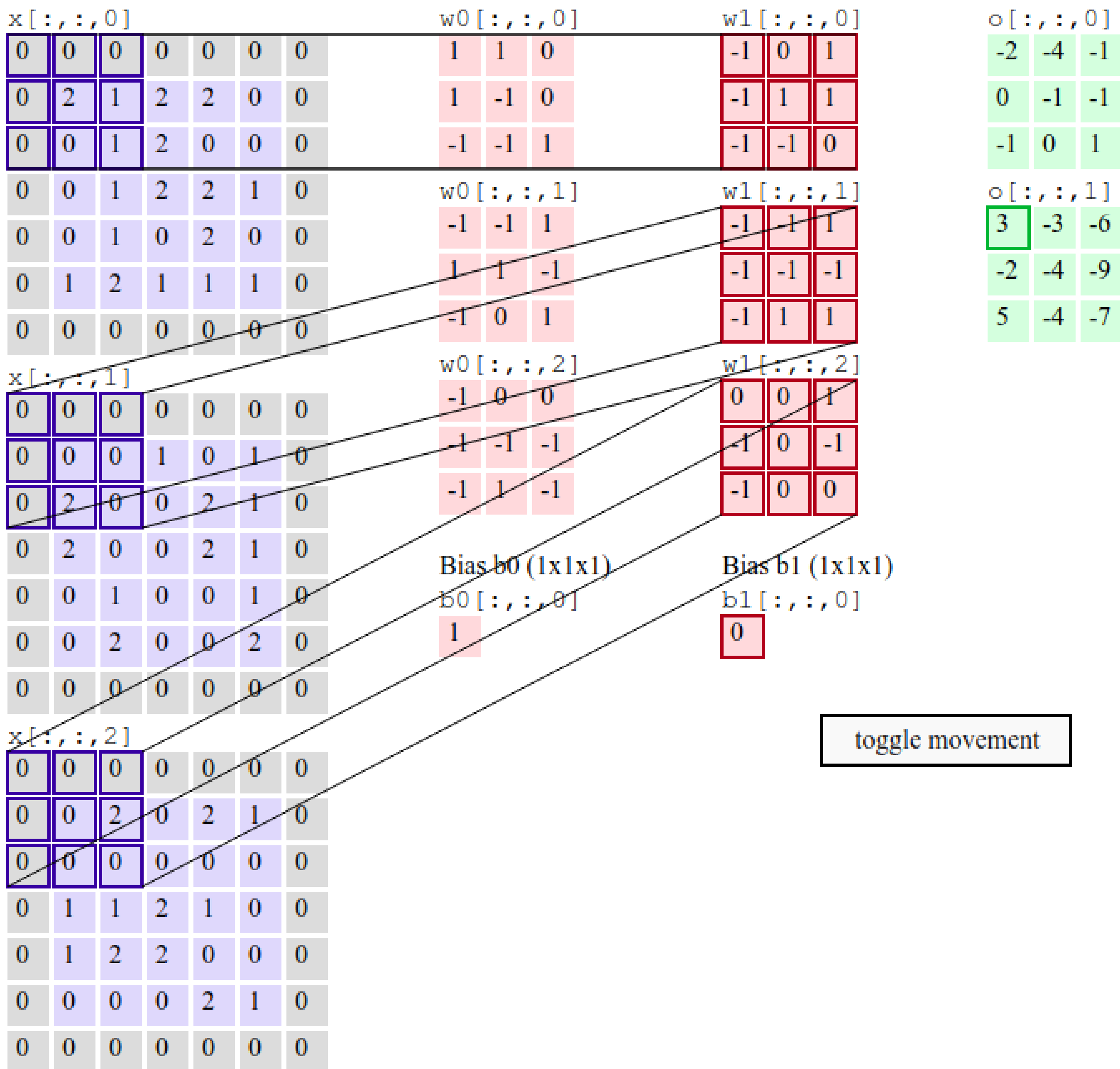

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement


x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

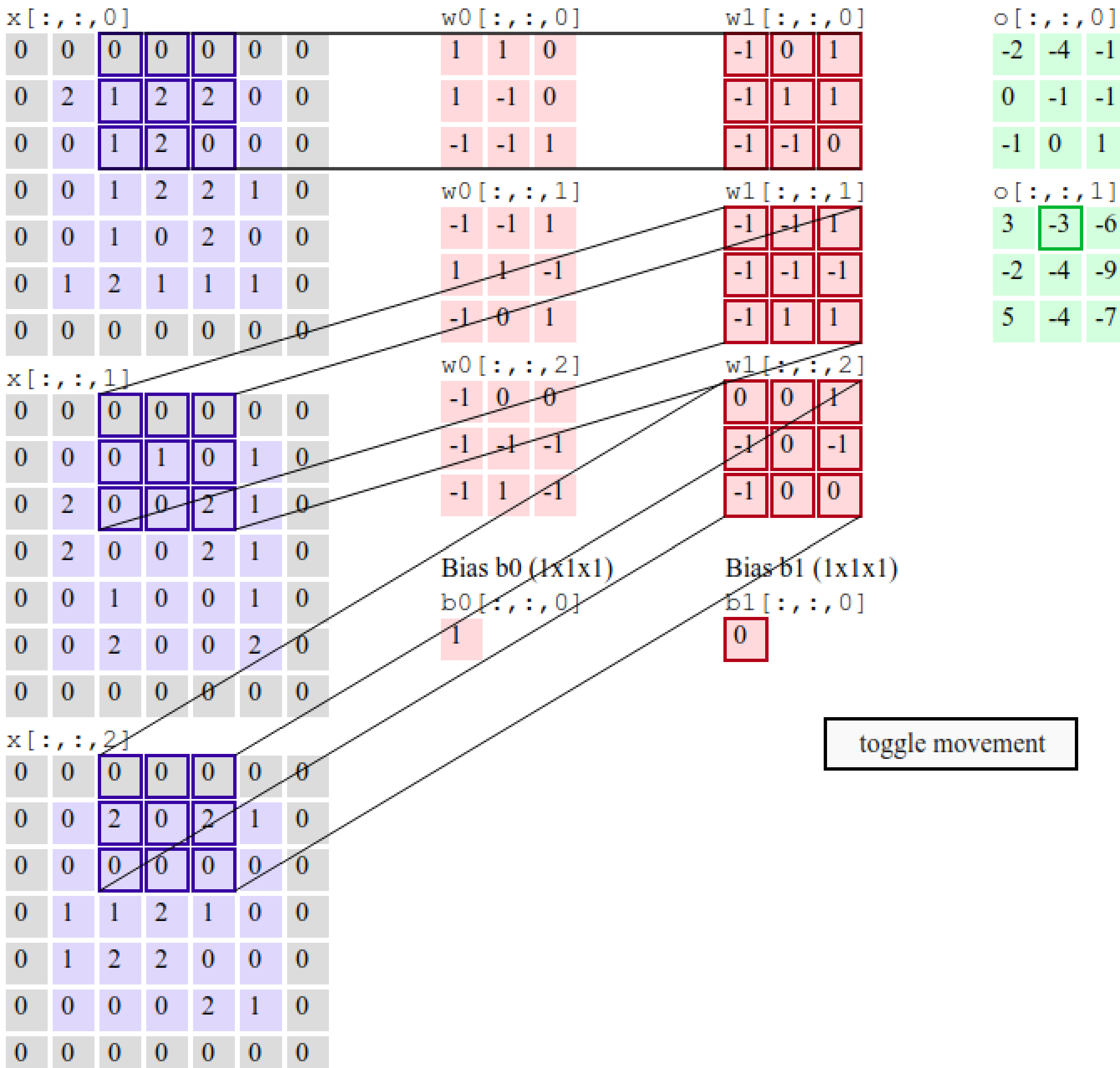

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

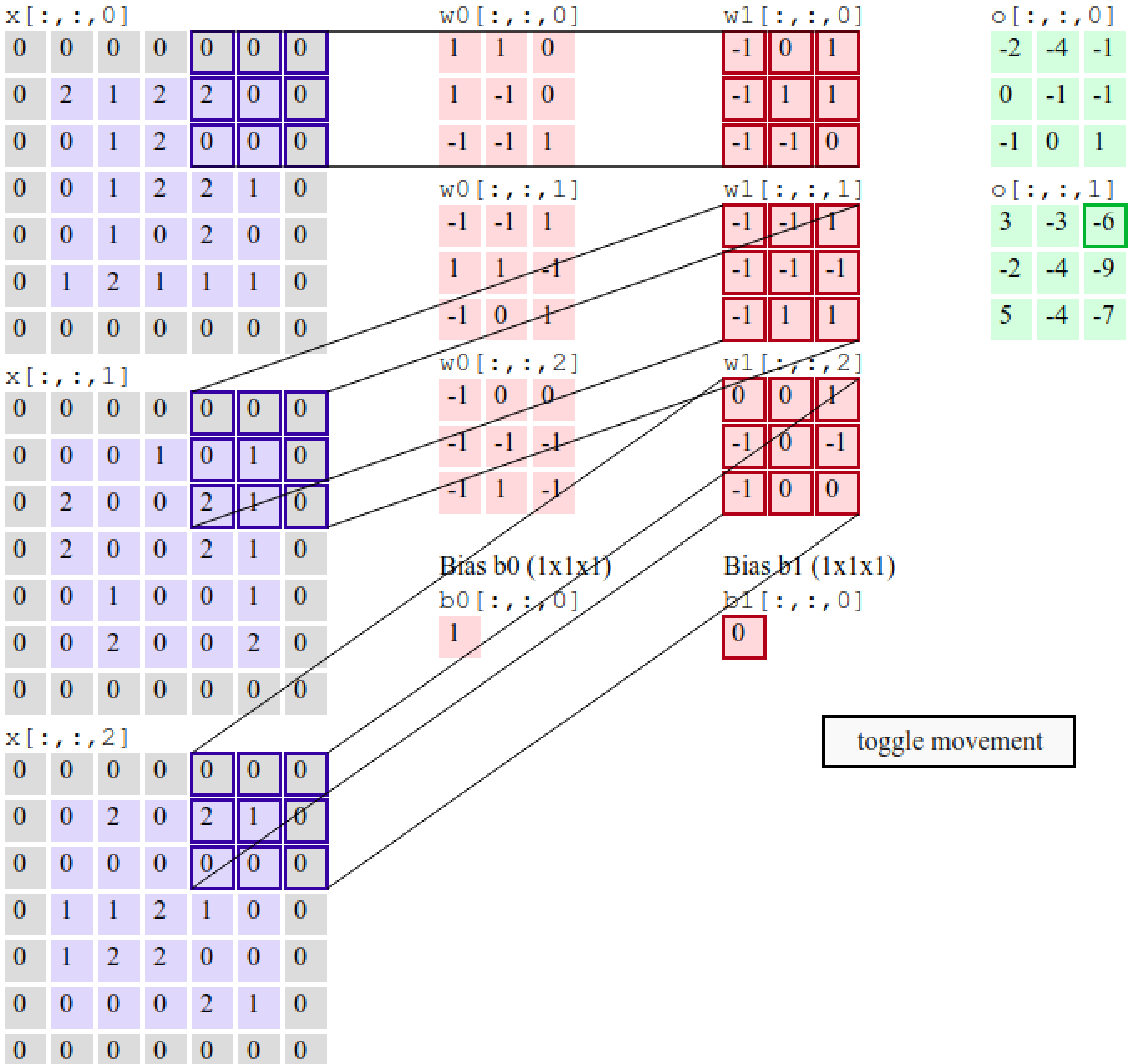

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

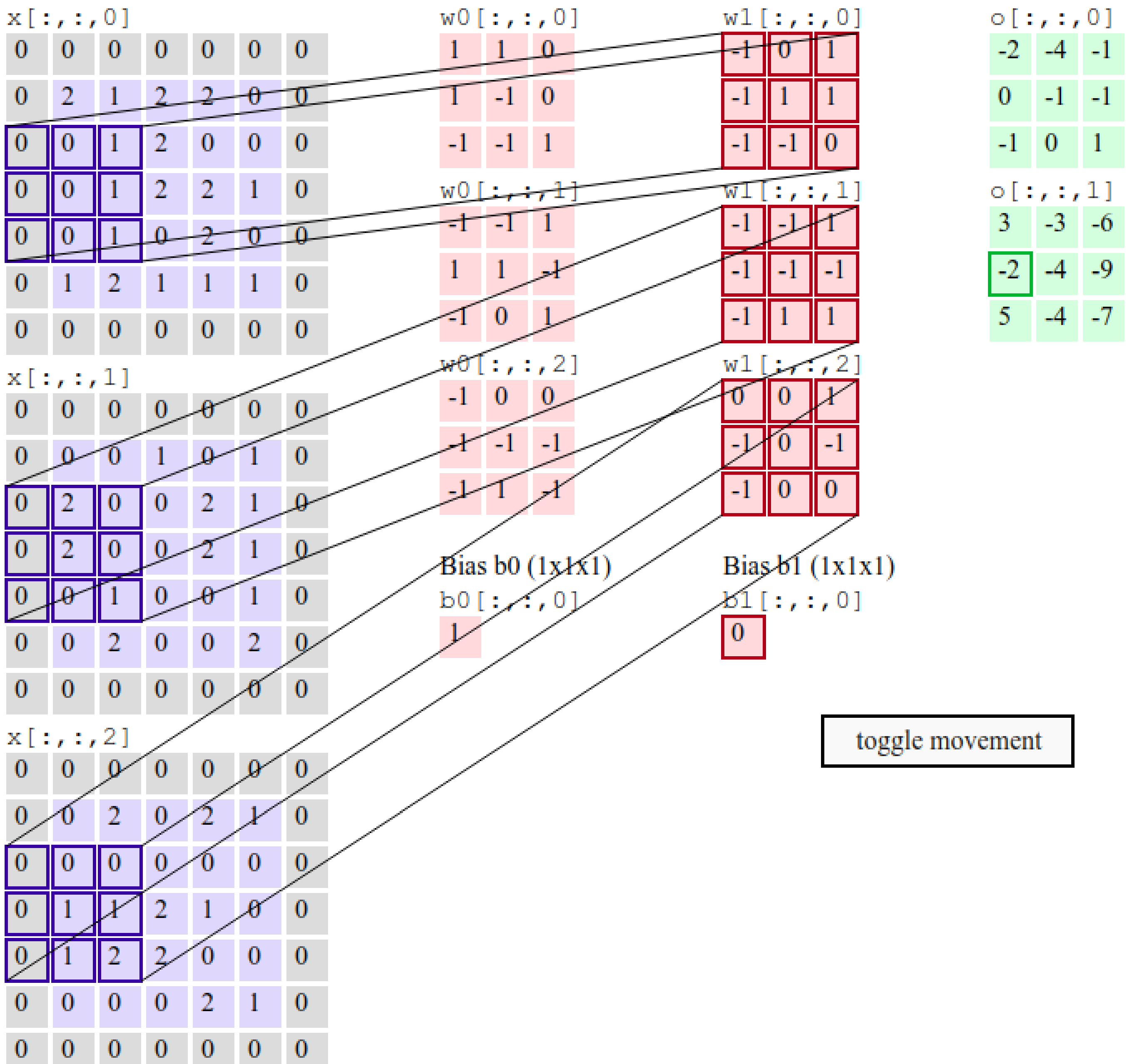

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

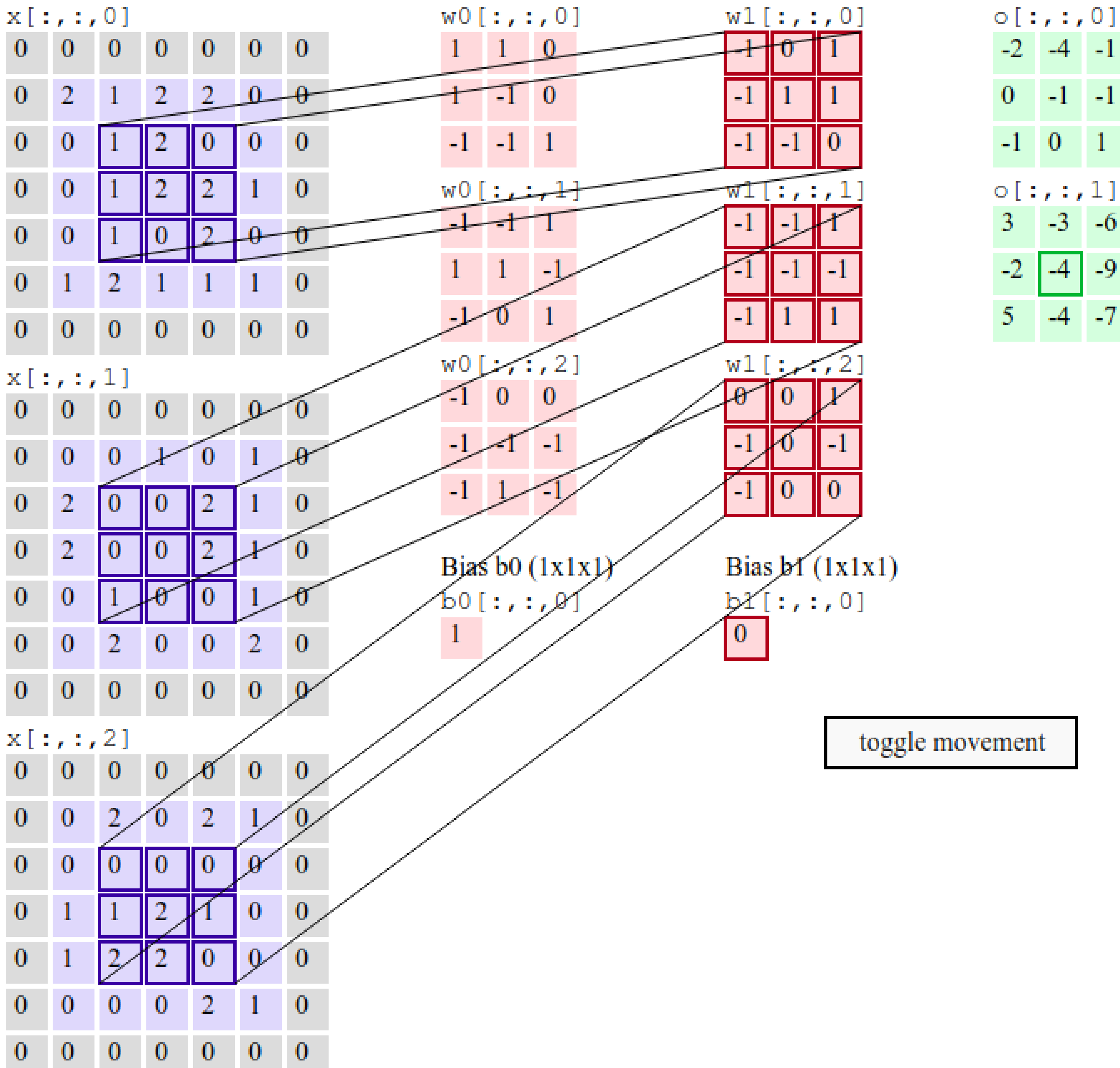


x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b0 (1x1x1)
b0[:,:,0]
1
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

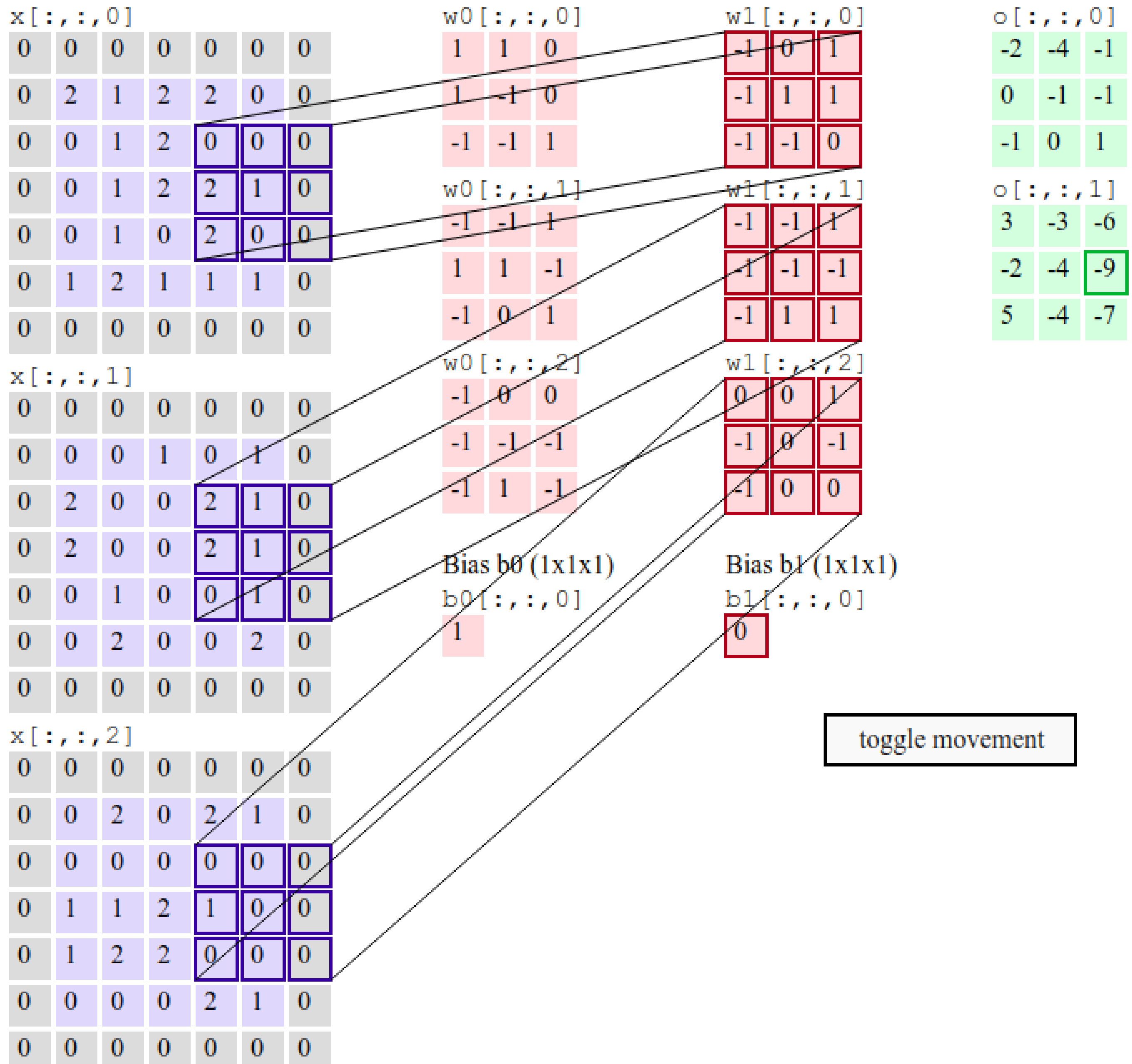

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement


x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

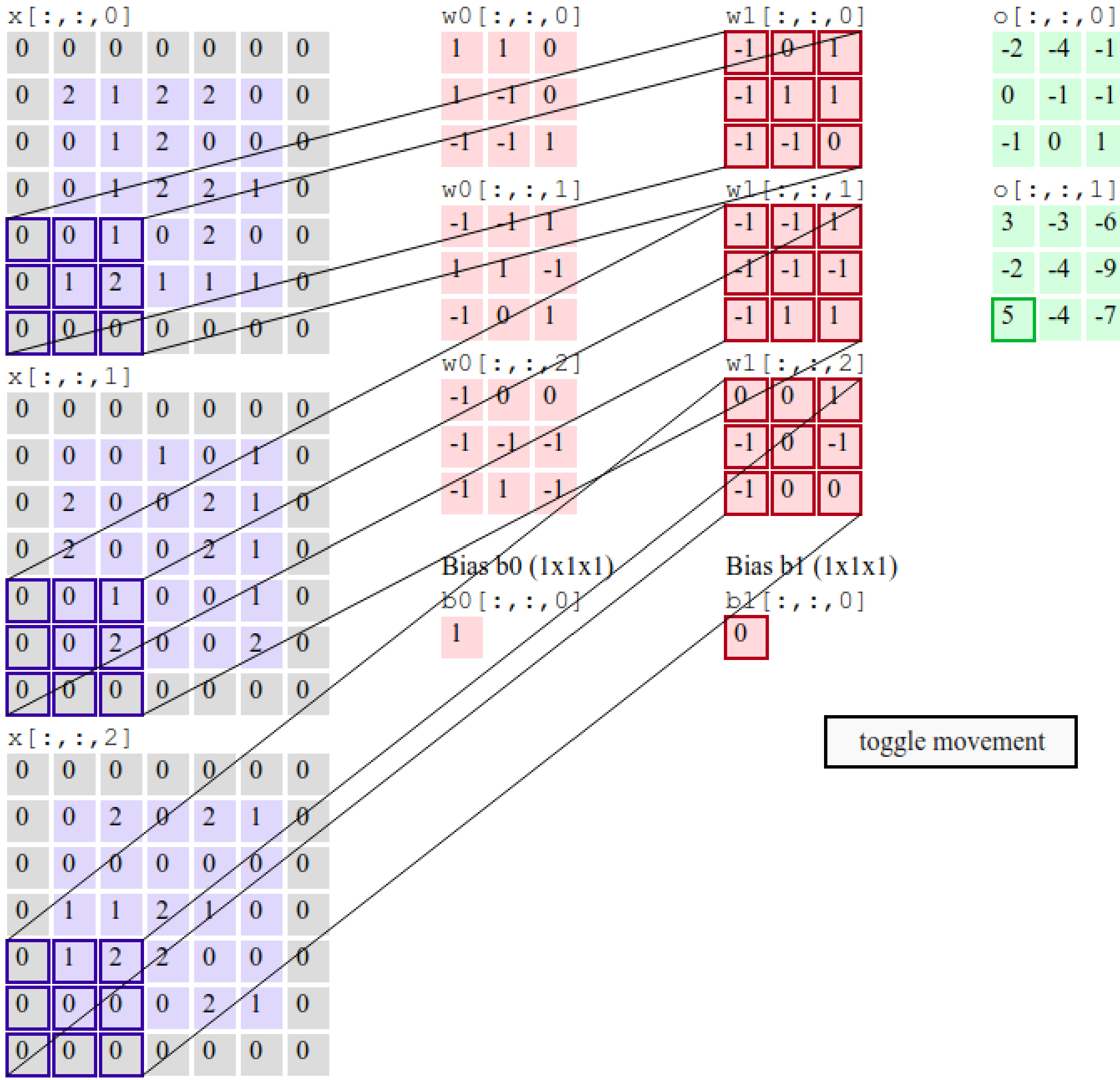

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

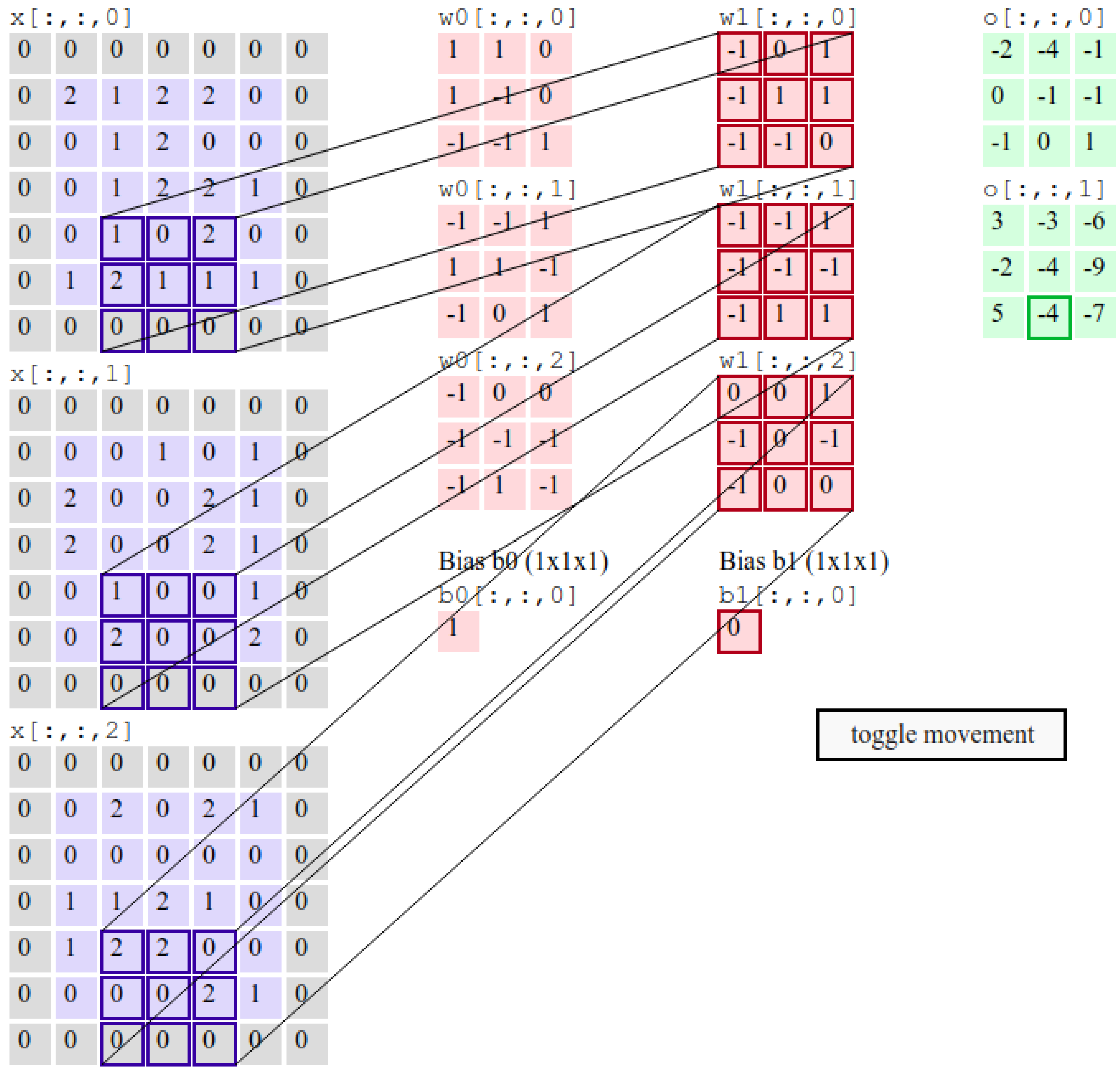


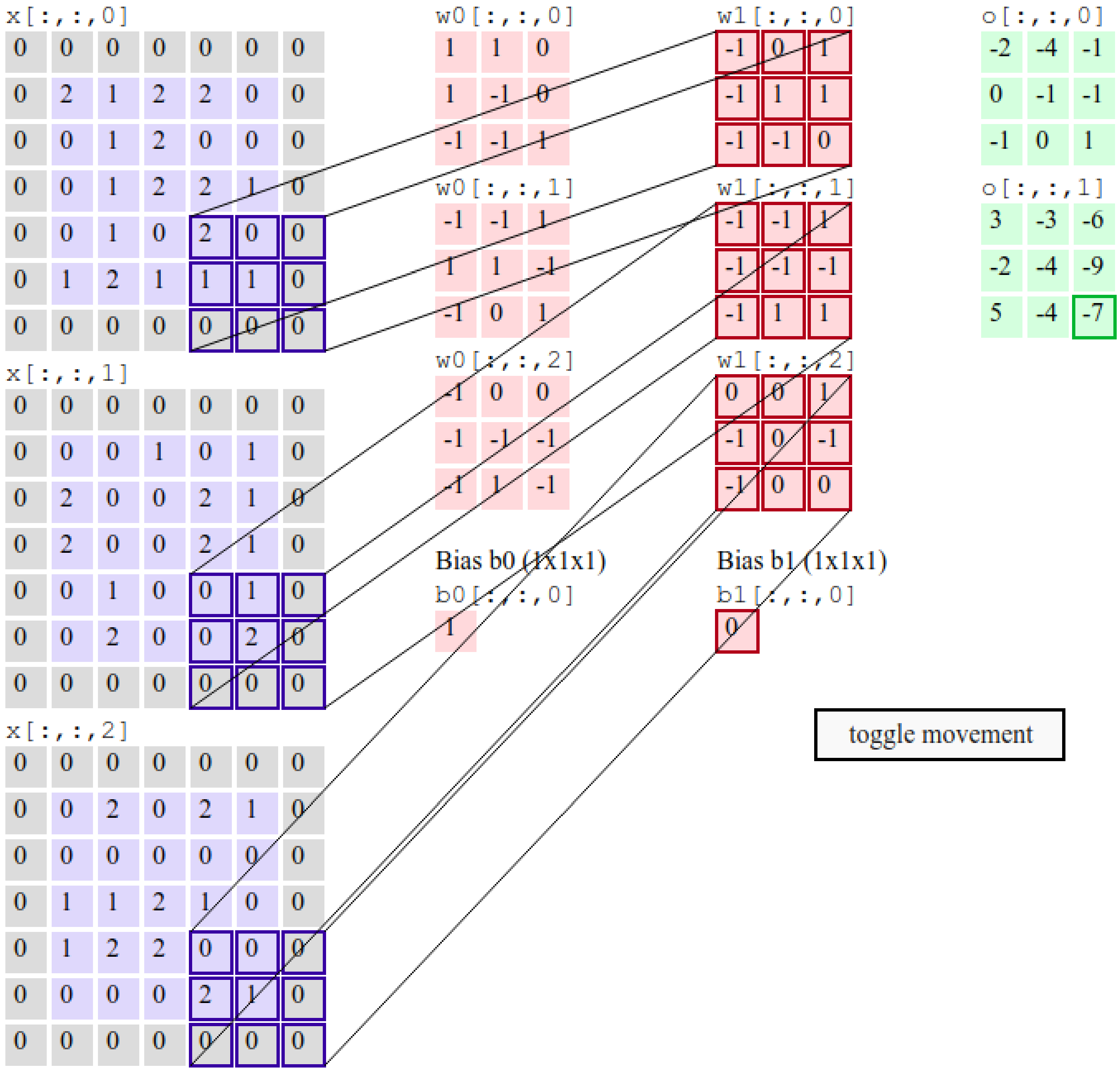
x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement



x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

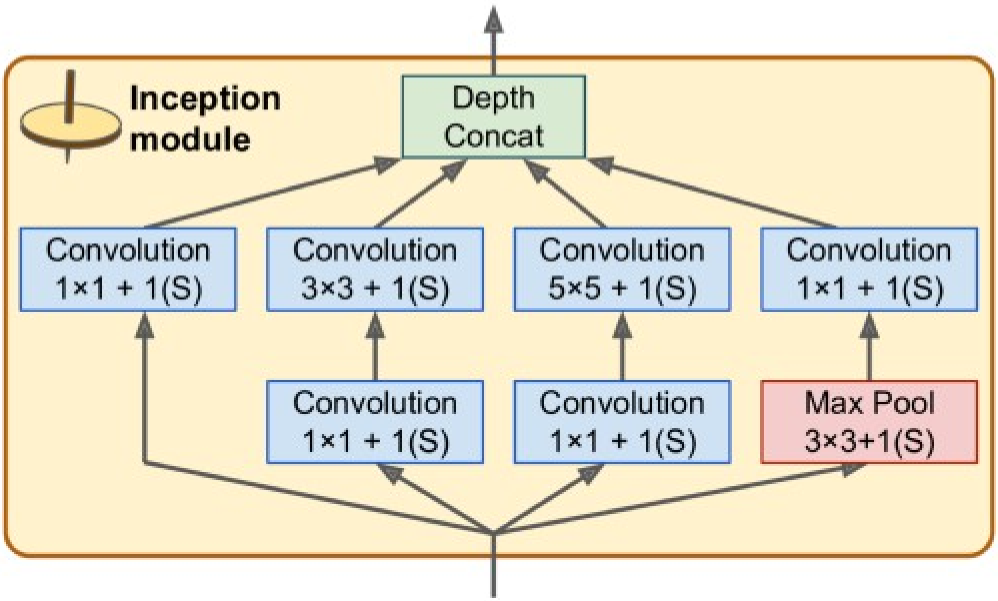

x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement


x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement


x[:,:,0]
0 0 0 0 0 0 0
0 2 1 2 2 0 0
0 0 1 2 0 0 0
0 0 1 2 2 1 0
0 0 1 0 2 0 0
0 1 2 1 1 1 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 0 1 0 1 0
0 2 0 0 2 1 0
0 2 0 0 2 1 0
0 0 1 0 0 1 0
0 0 2 0 0 2 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 0 2 0 2 1 0
0 0 0 0 0 0 0
0 1 1 2 1 0 0
0 1 2 2 0 0 0
0 0 0 0 2 1 0
0 0 0 0 0 0 0
w0[:,:,0]
1 1 0
1 -1 0
-1 -1 1
w0[:,:,1]
-1 -1 1
1 1 -1
-1 0 1
w0[:,:,2]
-1 0 0
-1 -1 -1
-1 1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
w1[:,:,0]
-1 0 1
-1 1 1
-1 -1 0
w1[:,:,1]
-1 -1 1
-1 -1 -1
-1 1 1
w1[:,:,2]
0 0 1
-1 0 -1
-1 0 0
Bias b1 (1x1x1)
b1[:,:,0]
0
o[:,:,0]
-2 -4 -1
0 -1 -1
-1 0 1
o[:,:,1]
3 -3 -6
-2 -4 -9
5 -4 -7
toggle movement

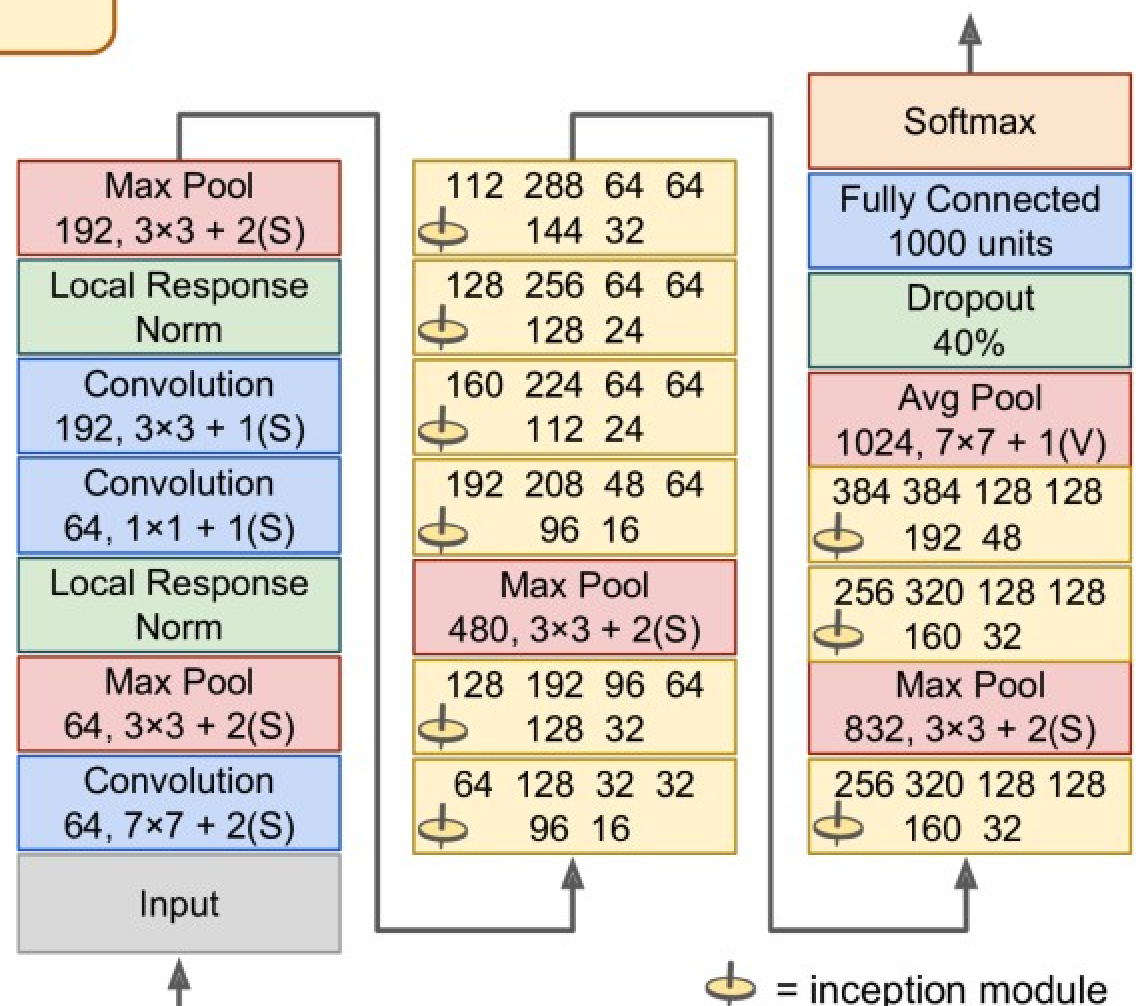
Two famous deep NN architecture

GoogleNet

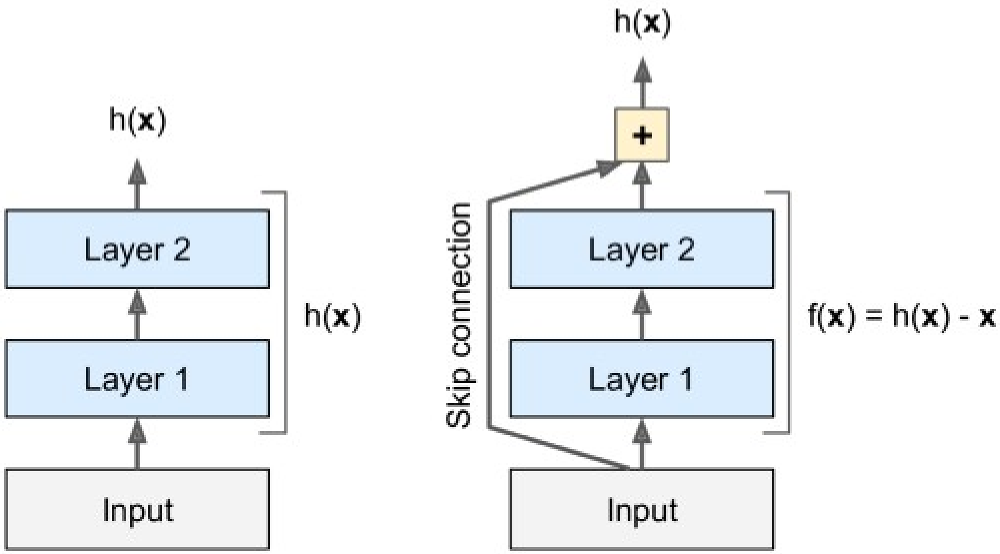


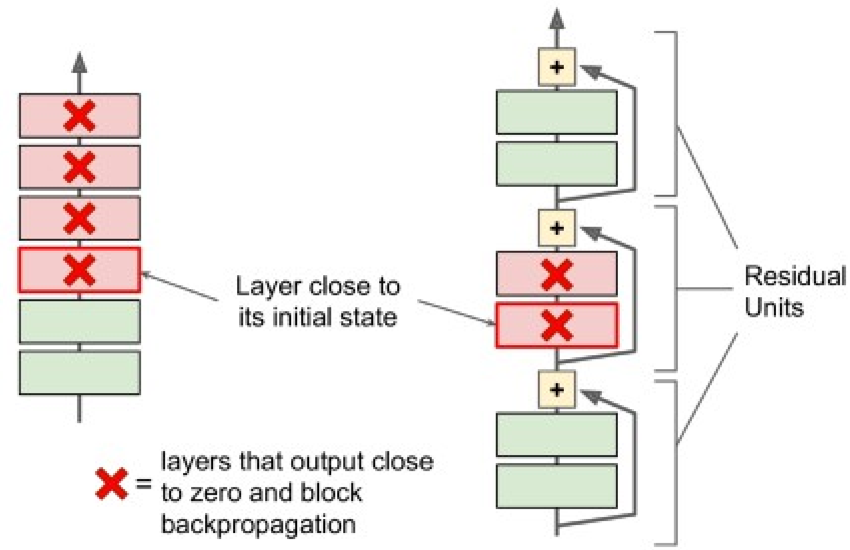
Developers – Christian Szegedy et al. from Google Research.

It won the ILSVRC 2014 challenge by pushing the top-5 error rate below 7%

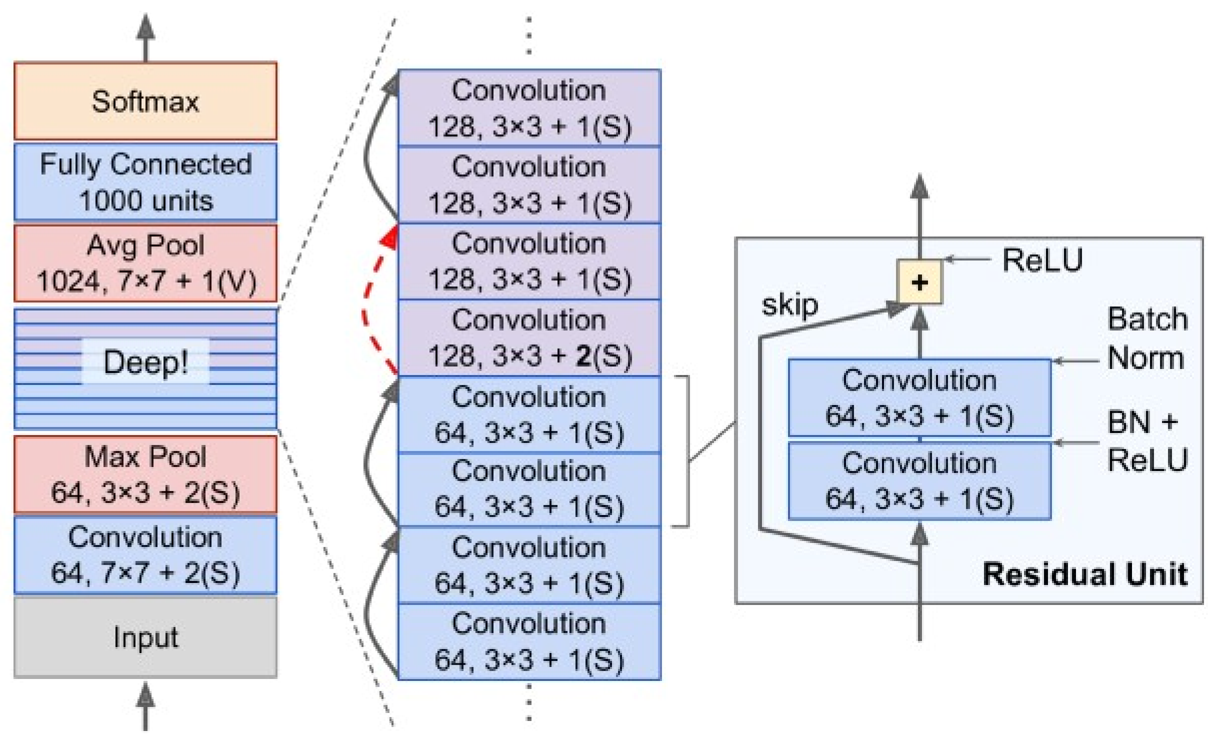


ResNet





winner of the ILSVRC 2015 challenge was the Residual Network (or ResNet), developed by Kaiming He et al.




That's all Folks!