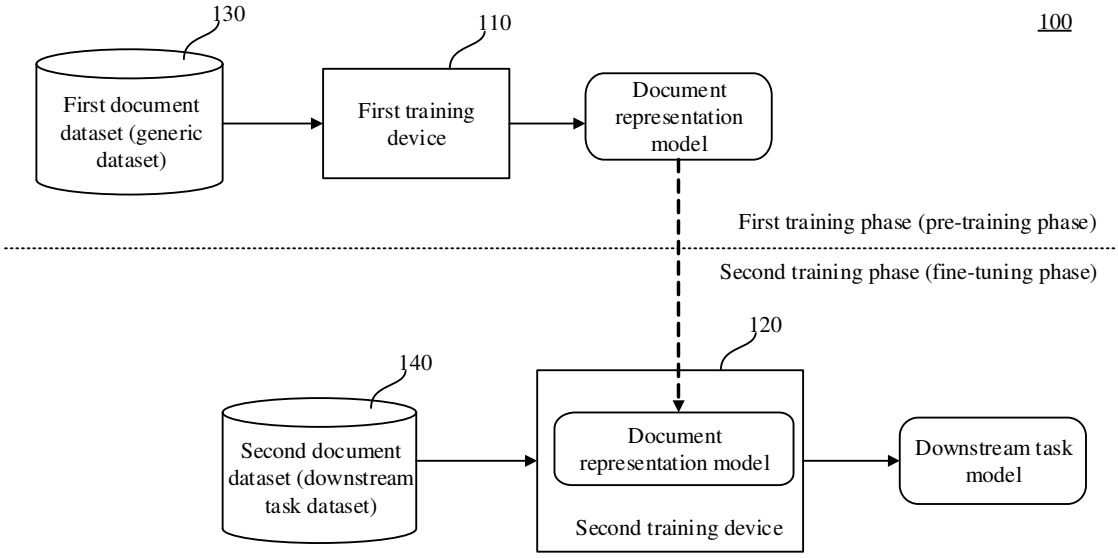

130
110
100
First document dataset (generic dataset)
First training device
Document representation model
First training phase (pre-training phase)
Second training phase (fine-tuning phase)
120
140
Second document dataset (downstream task dataset)
Document representation model
Second training device
Downstream task model

FIG. 1

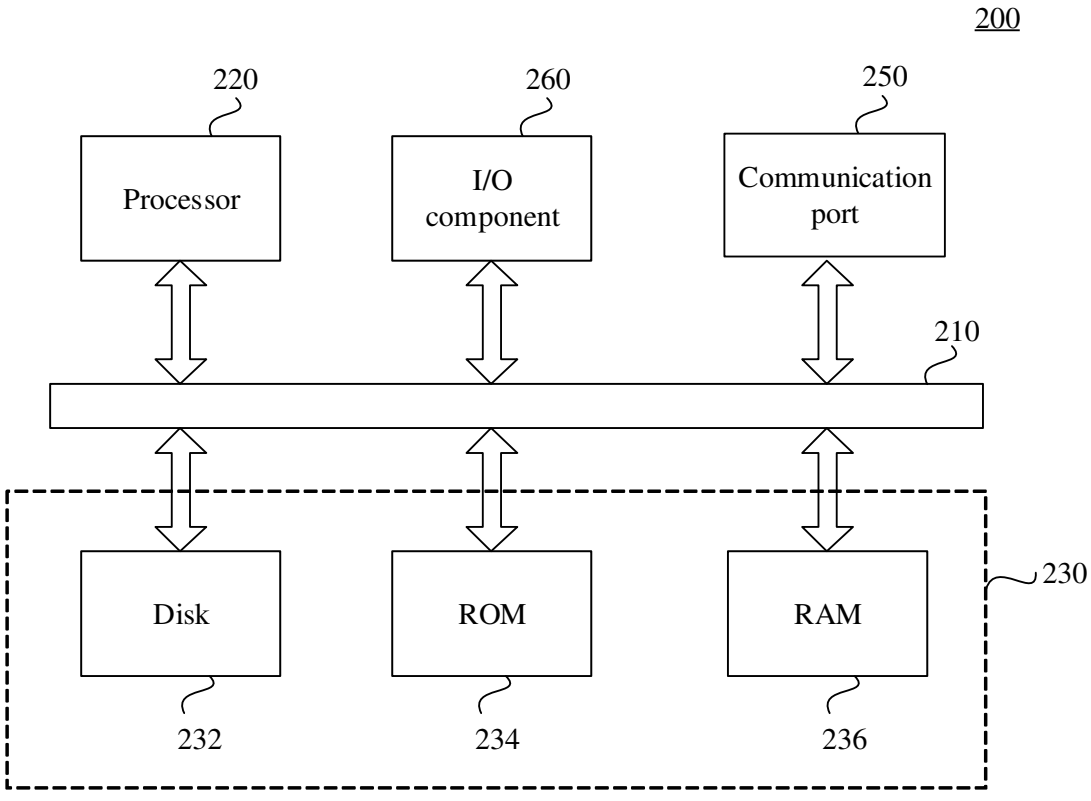

200
220
260
250
Processor
I/O component
Communication port
210
230
Disk
ROM
RAM
232
234
236

FIG. 2

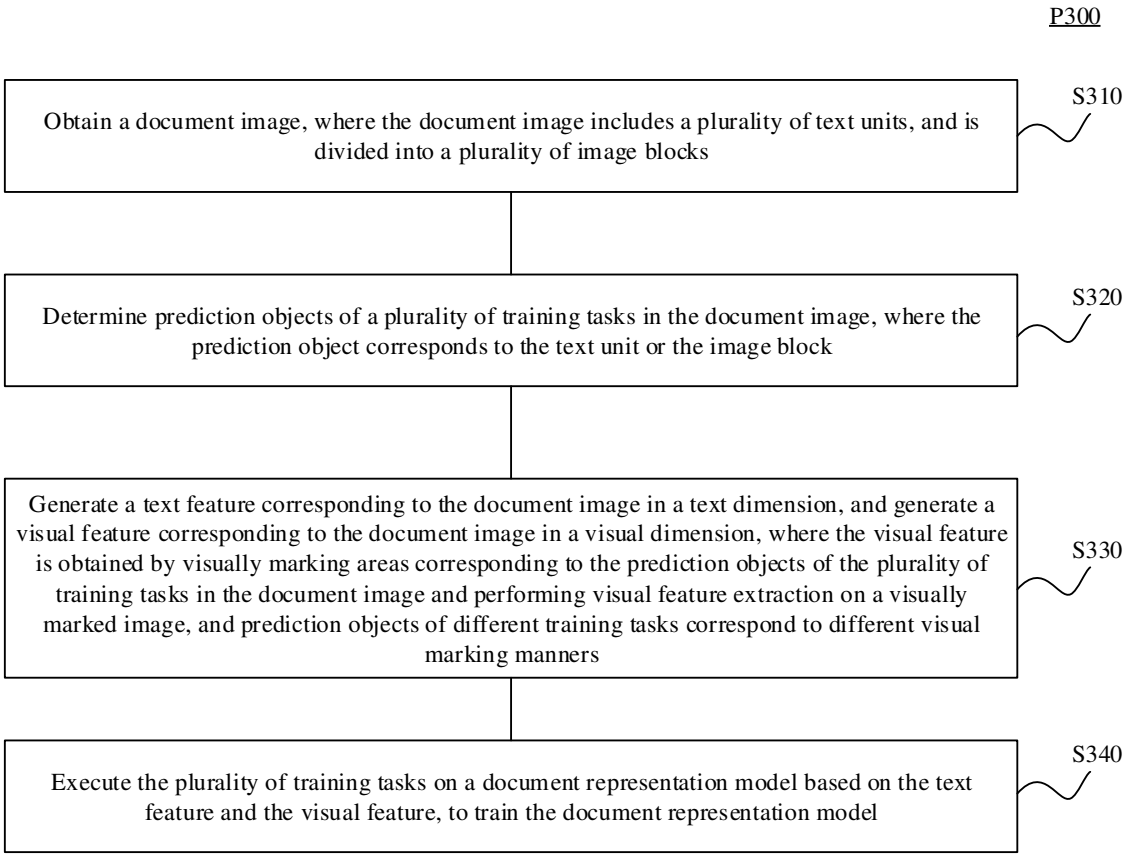
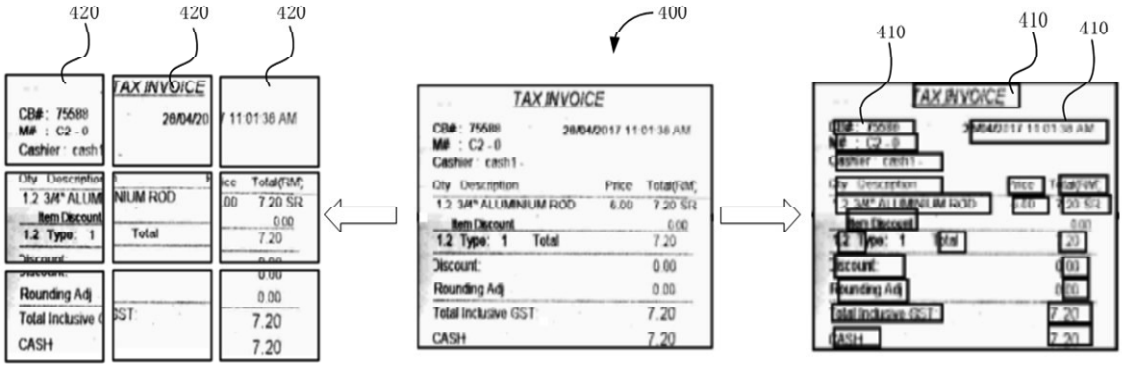


FIG. 3



(c) A plurality of image blocks

(a) Document image

(b) A plurality of text units

FIG. 4

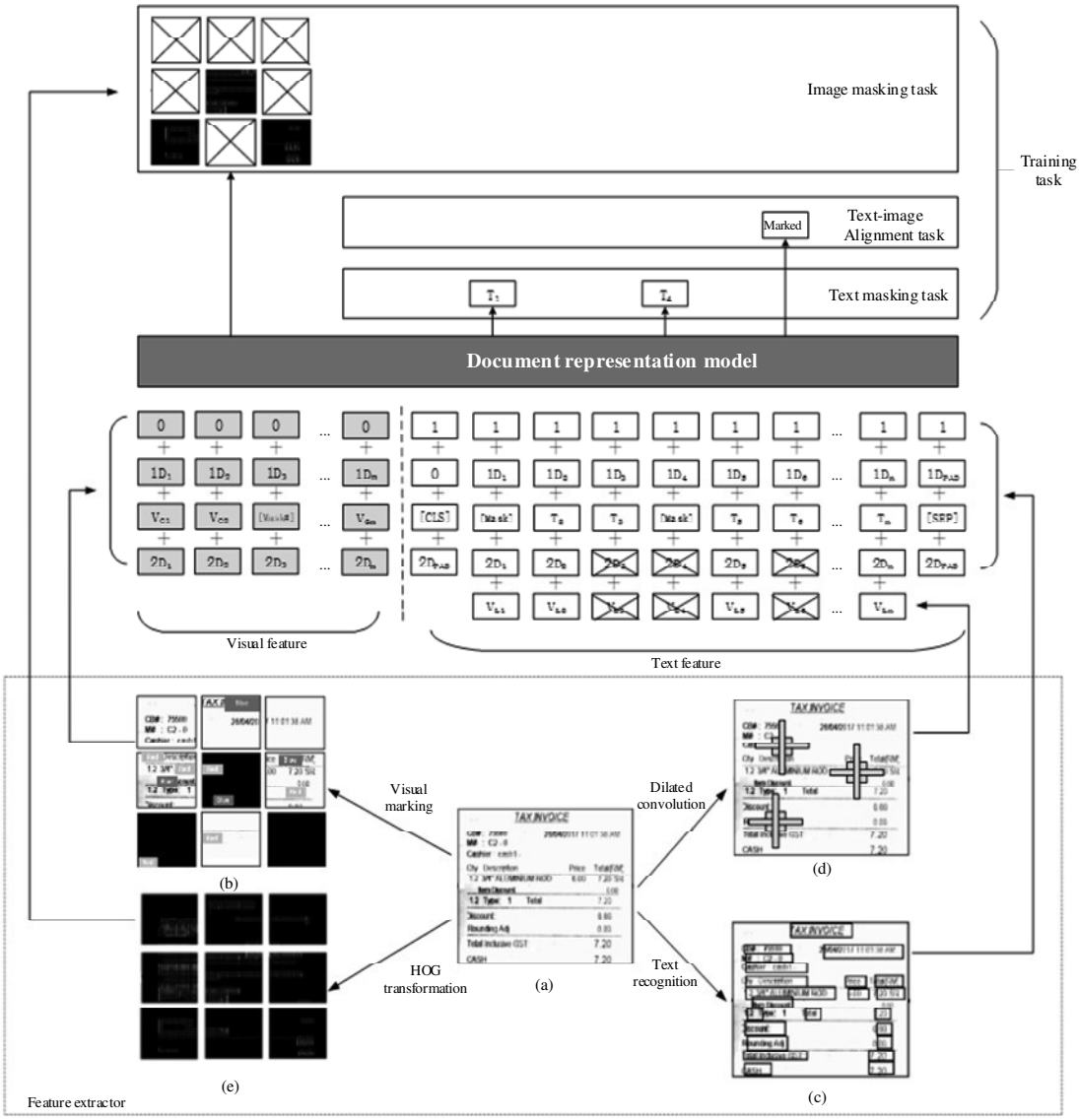


Image masking task
Training task
Marked
Text-image Alignment task
Text masking task
Document representation model
Visual feature
Text feature
Visual marking
Dilated convolution
HOG transformation
Text recognition
(a)
(b)
(c)
(d)
(e)
Feature extractor

FIG. 5

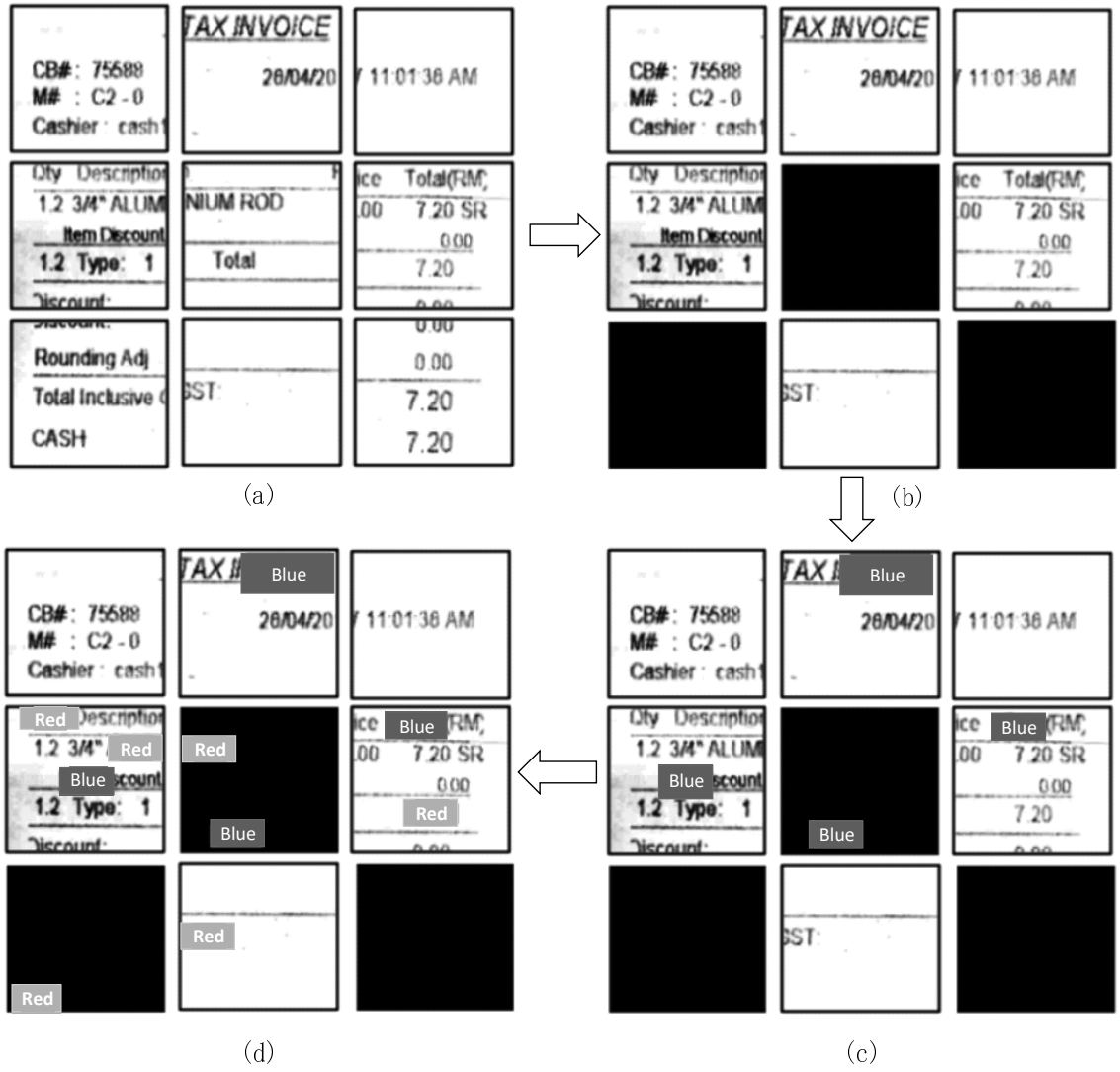

TAX INVOICE
CB#: 75588
M#: C2-0
Cashier: cash1
26/04/20
11:01:36 AM
Qty Description
1.2 3/4" ALUM
Item Discount
1.2 Type: 1
Total
Rounding Adj
Total Inclusive
CASH
7.20 SR
0.00
7.20
(a)
(b)
Blue
Red
(c)
(d)

FIG. 6

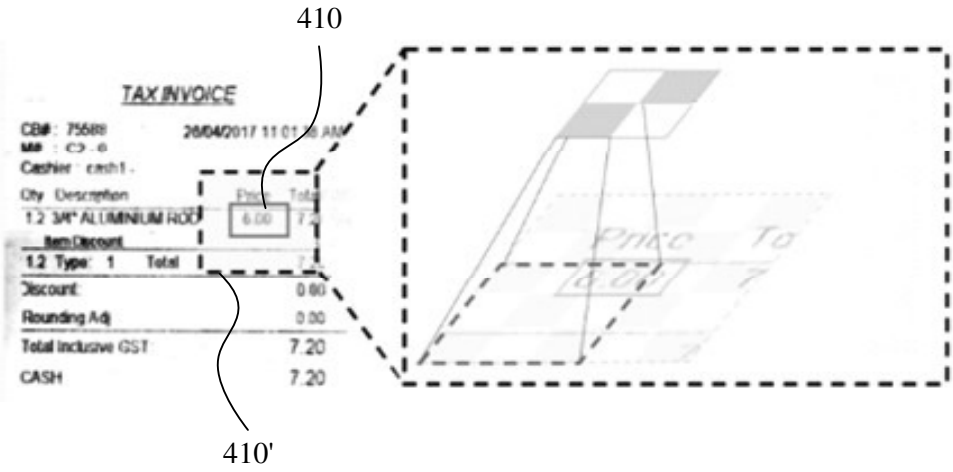

410
TAX INVOICE
Cashier : cash1 -
Qty Description
1.2 3/4" ALUMINIUM ROD
Item Discount
1.2 Type: 1 Total
Discount
0.00
Rounding Adj
0.00
Total Inclusive GST
7.20
CASH
7.20
410'

FIG. 7

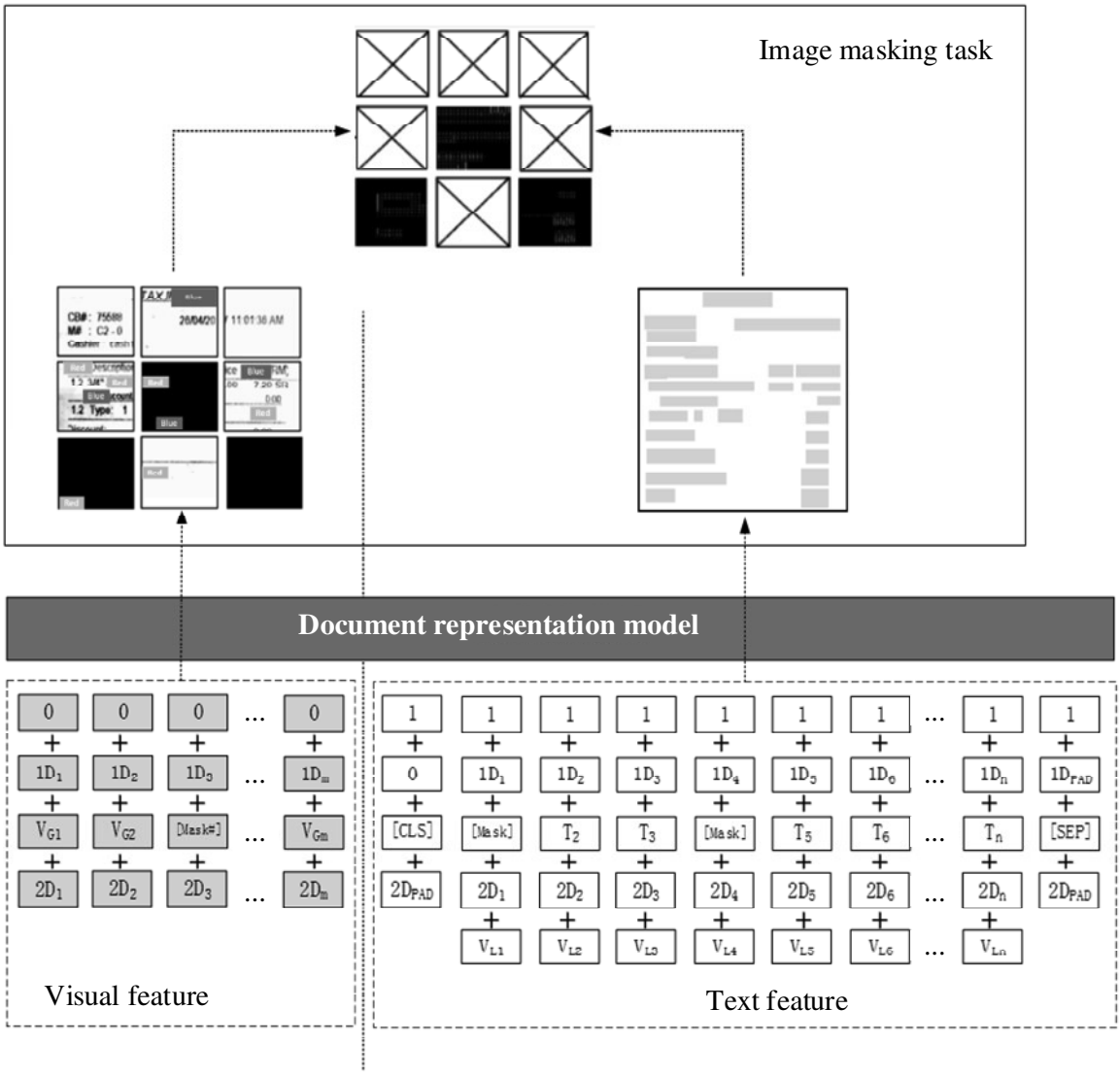


Image masking task
Document representation model
Visual feature
Text feature

FIG. 8