

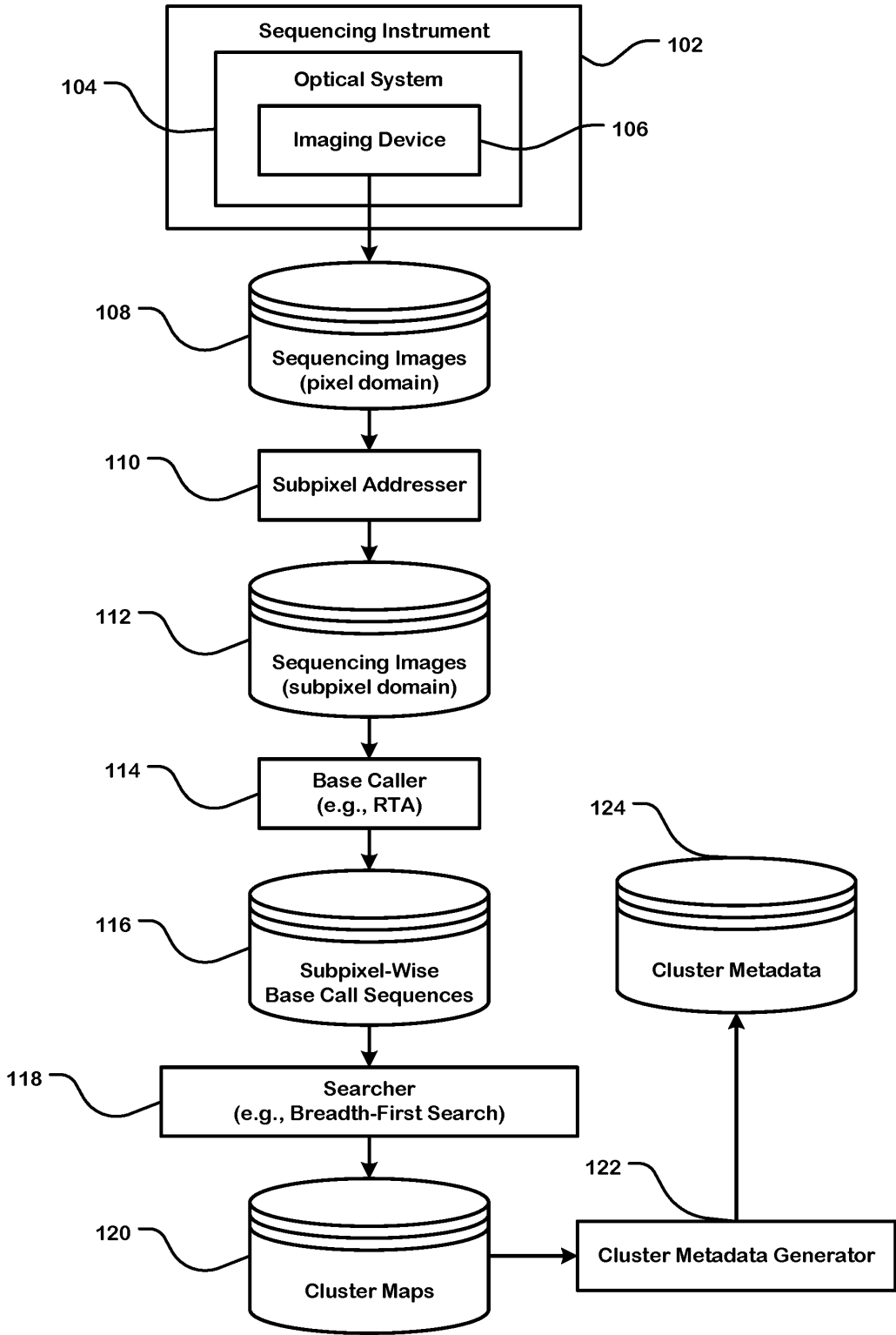


Figure 1

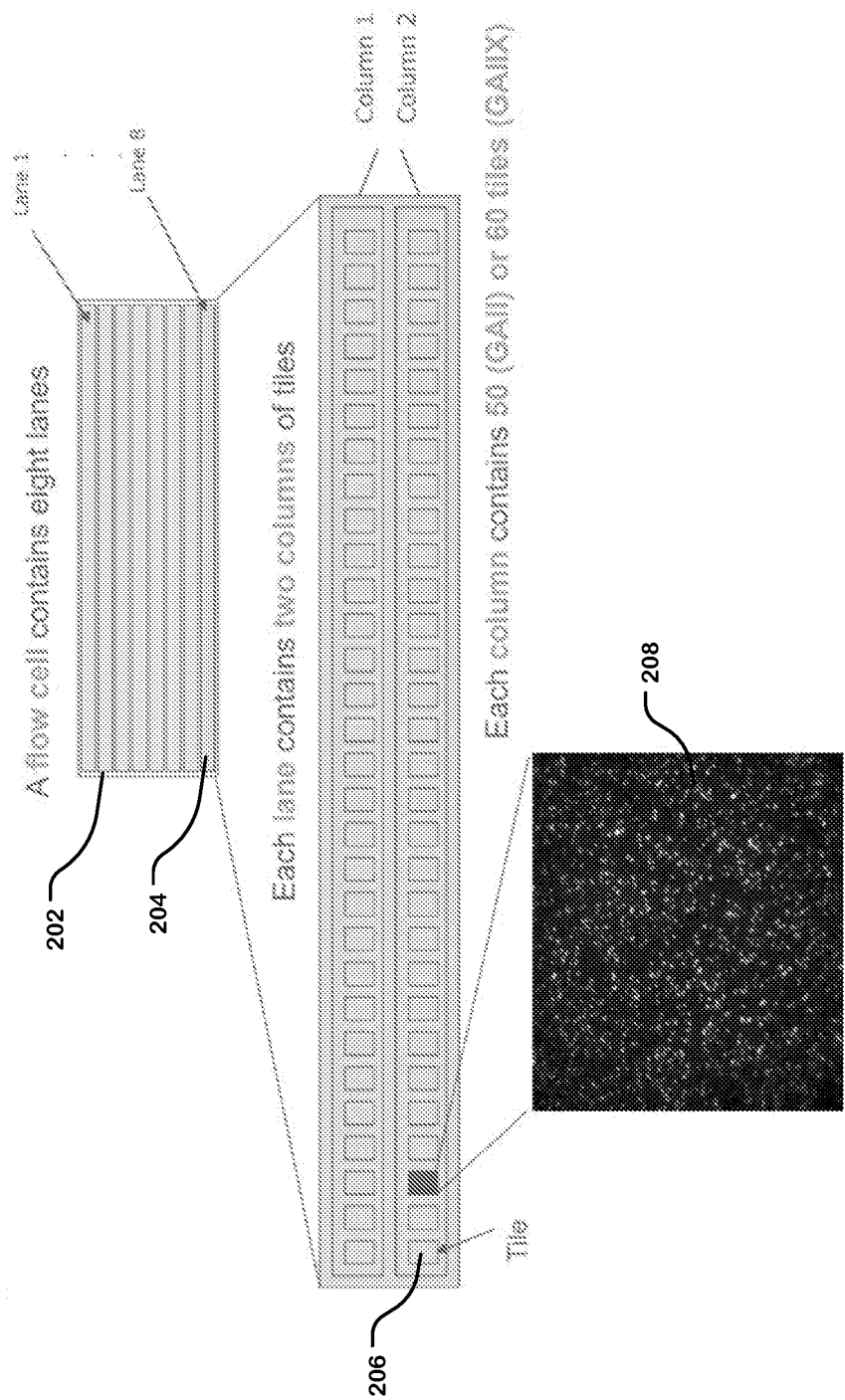


Figure 2

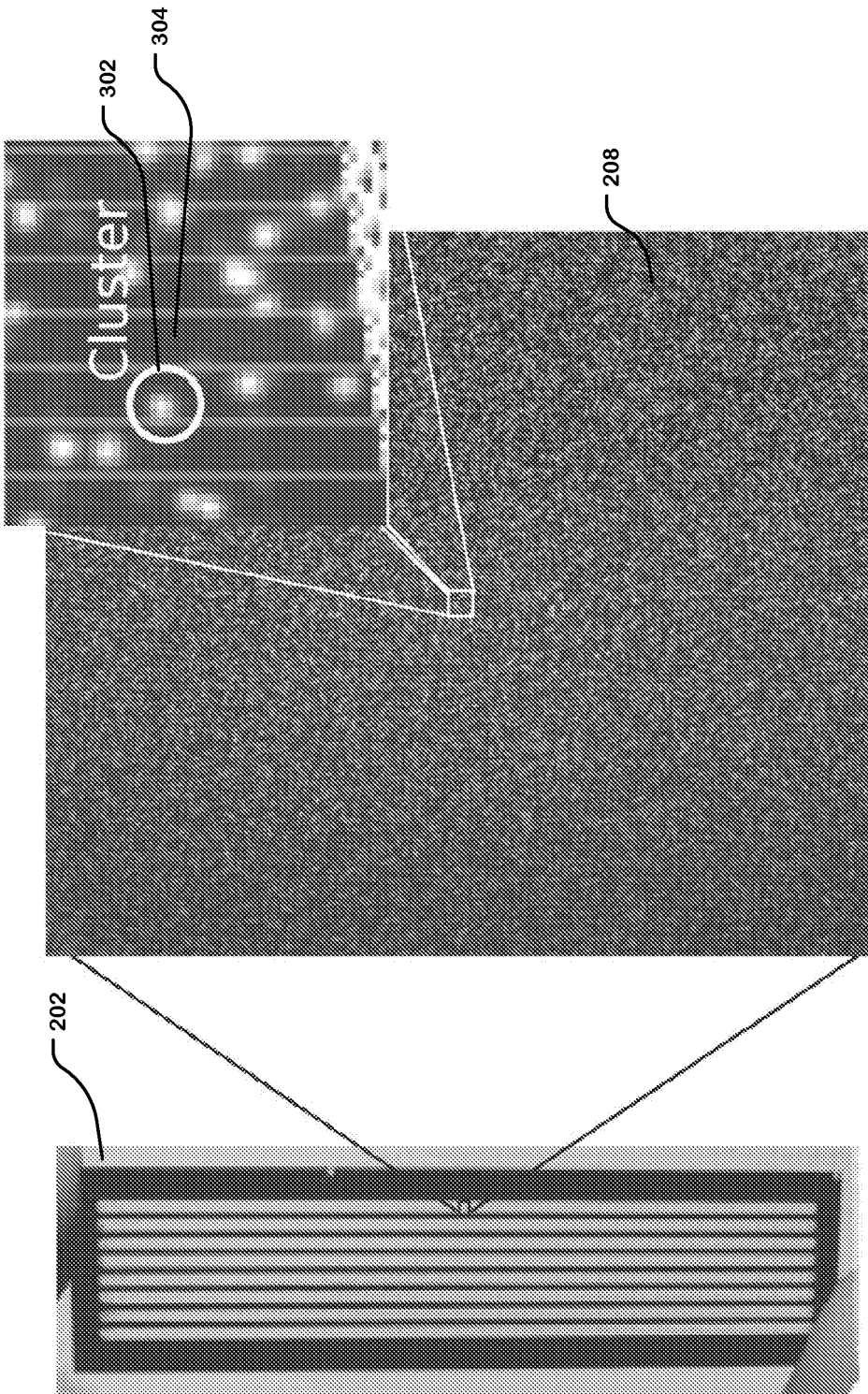


Figure 3

Sequencing Images in Pixel Domain 108

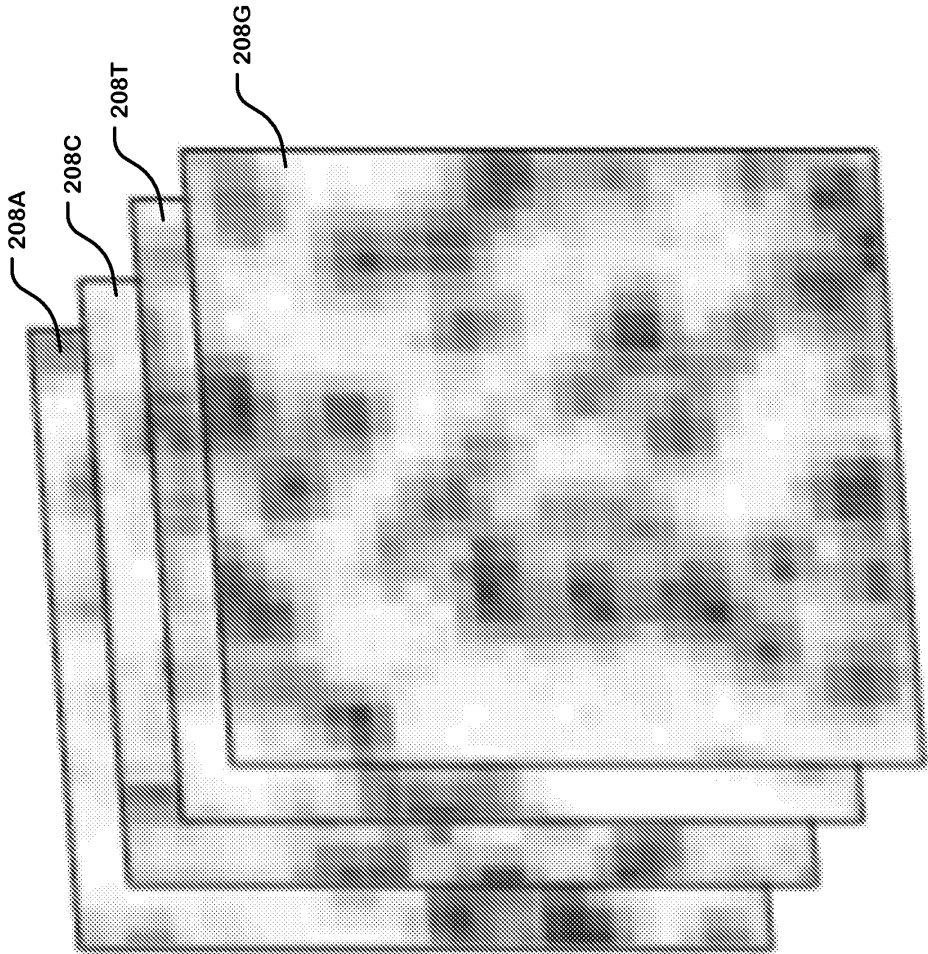


Figure 4

Subpixels to be Base Called

502

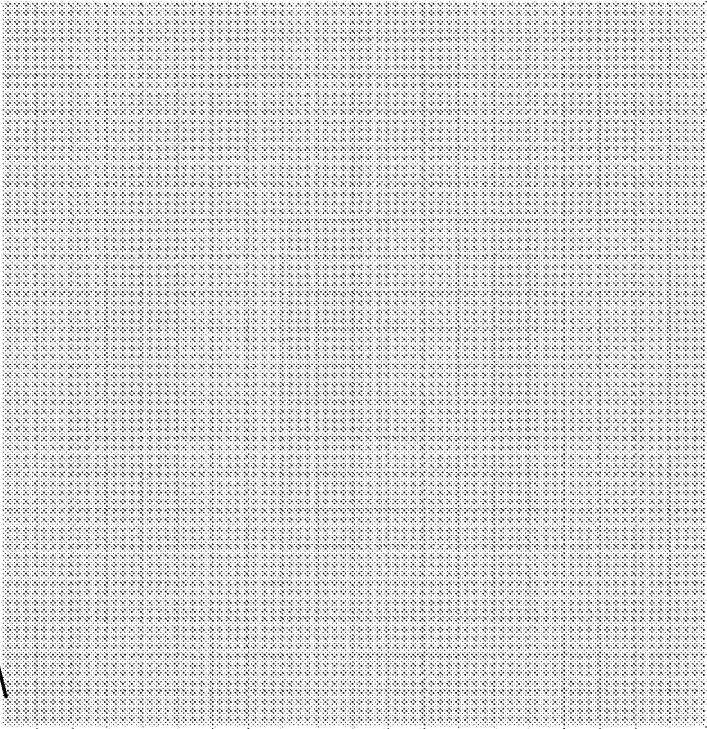
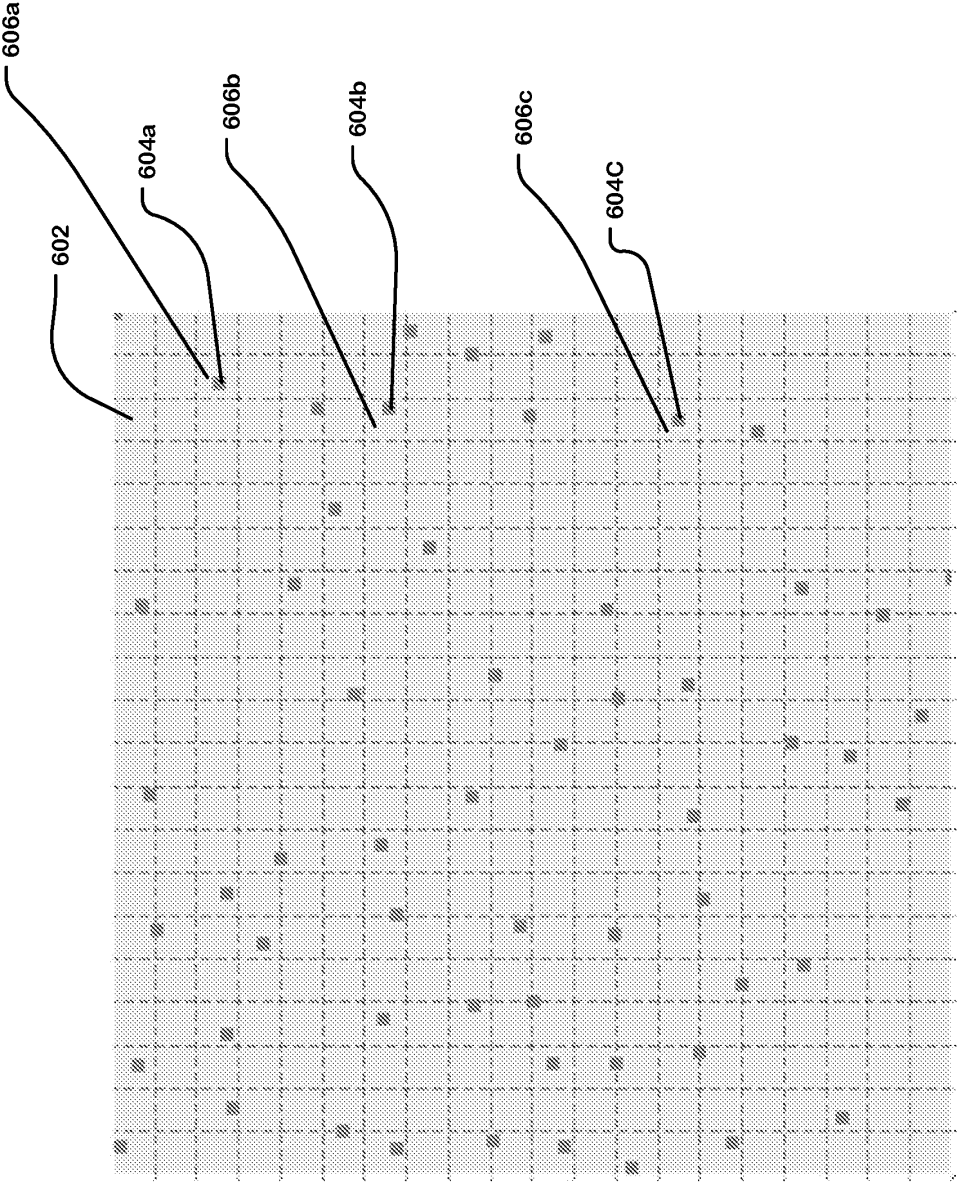


Figure 5



Cluster Centers (RTA) and Origin Subpixels

Figure 6

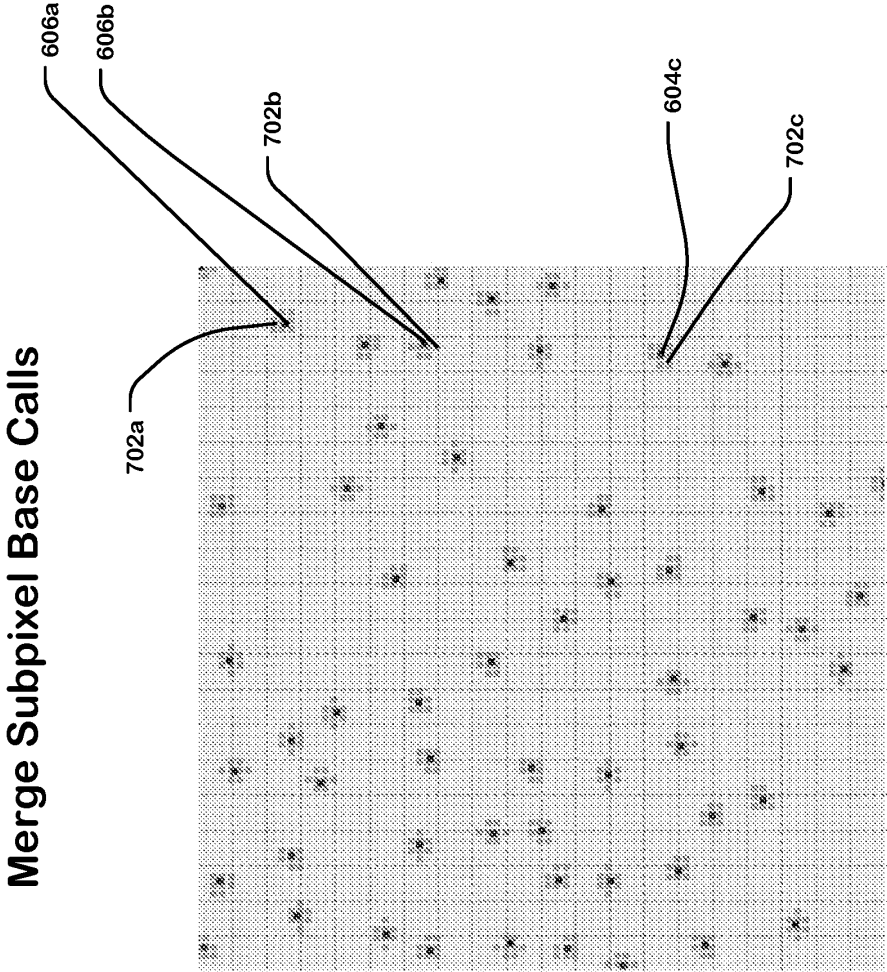


Figure 7

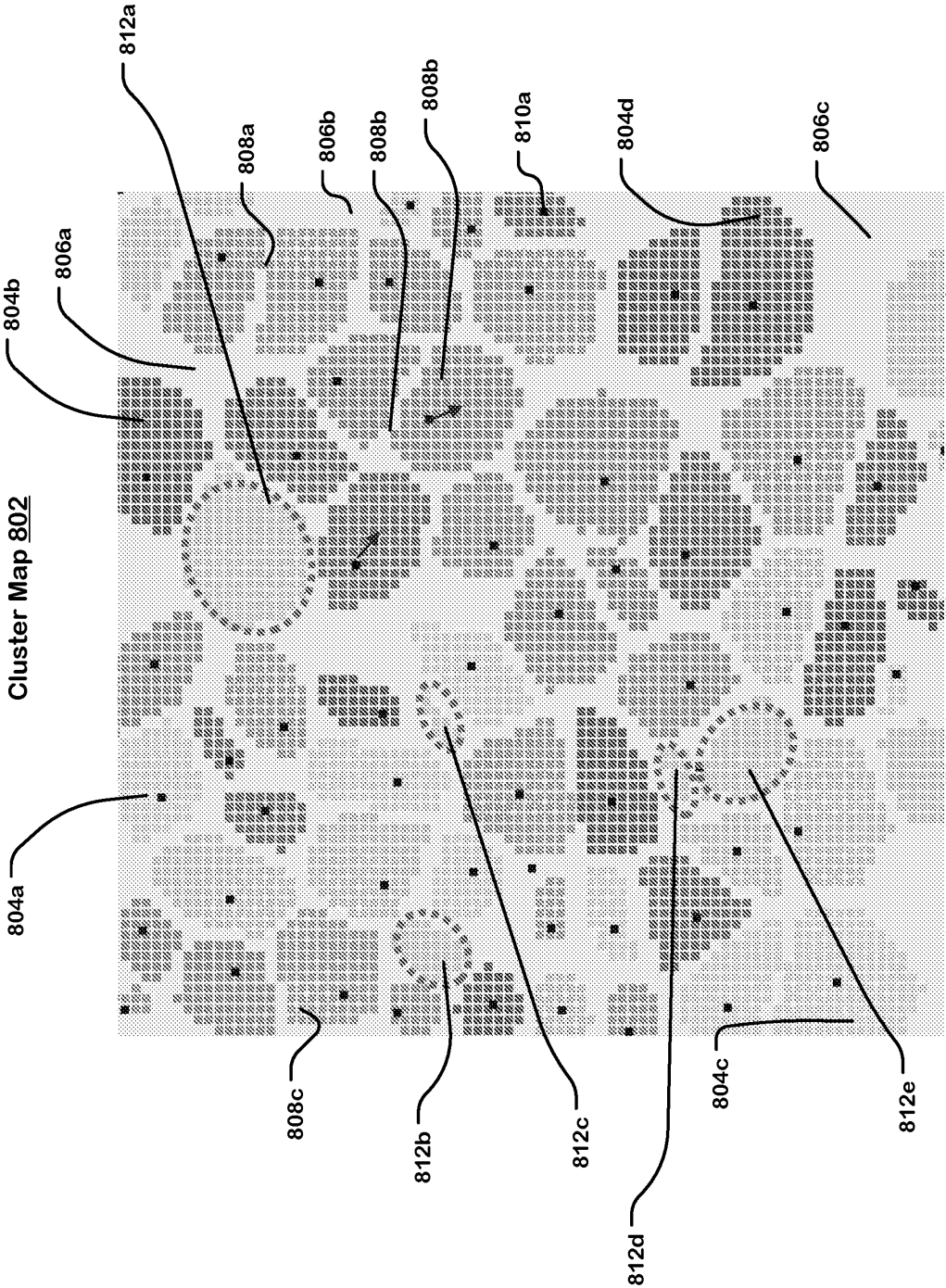


Figure 8a

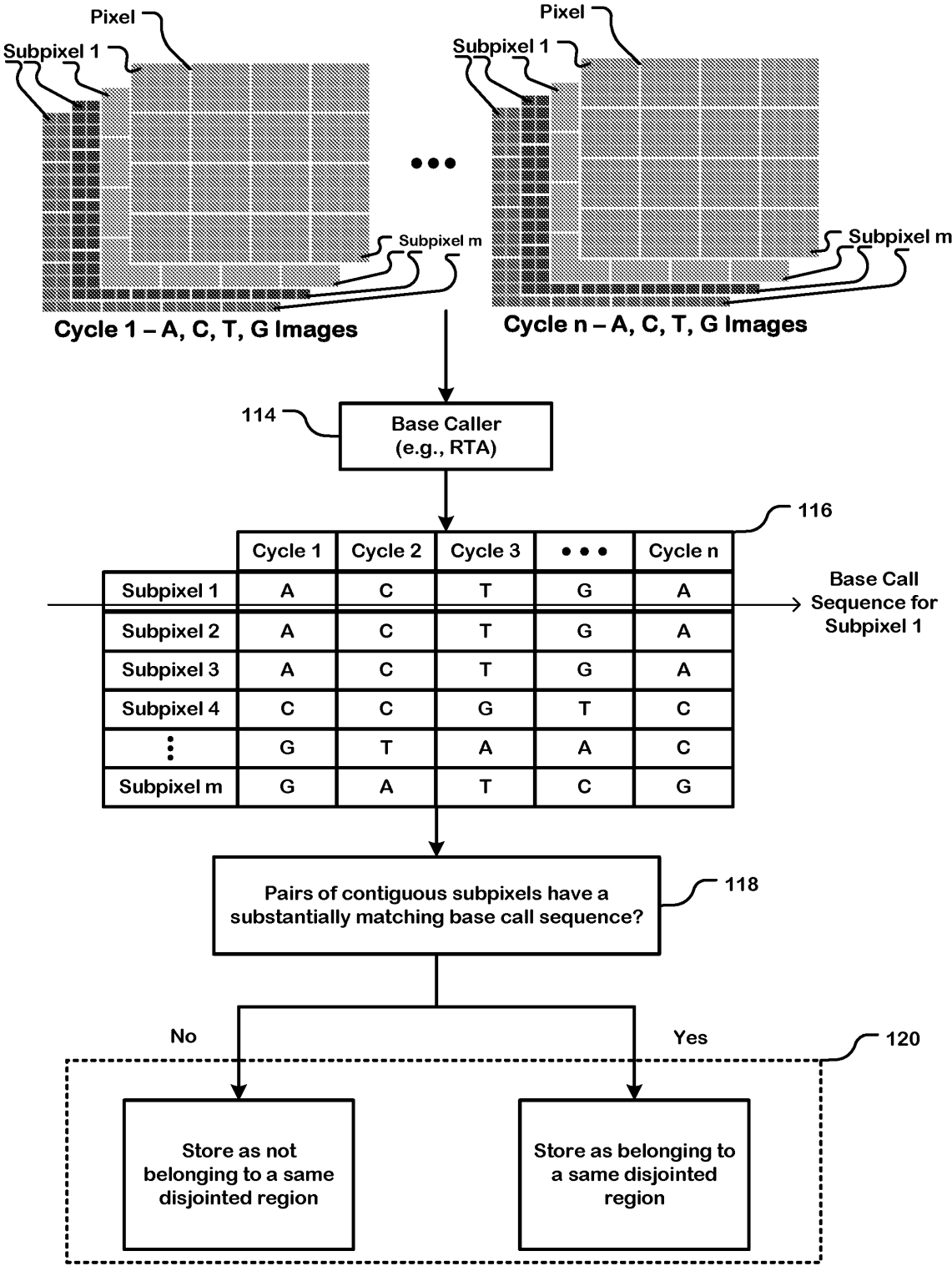
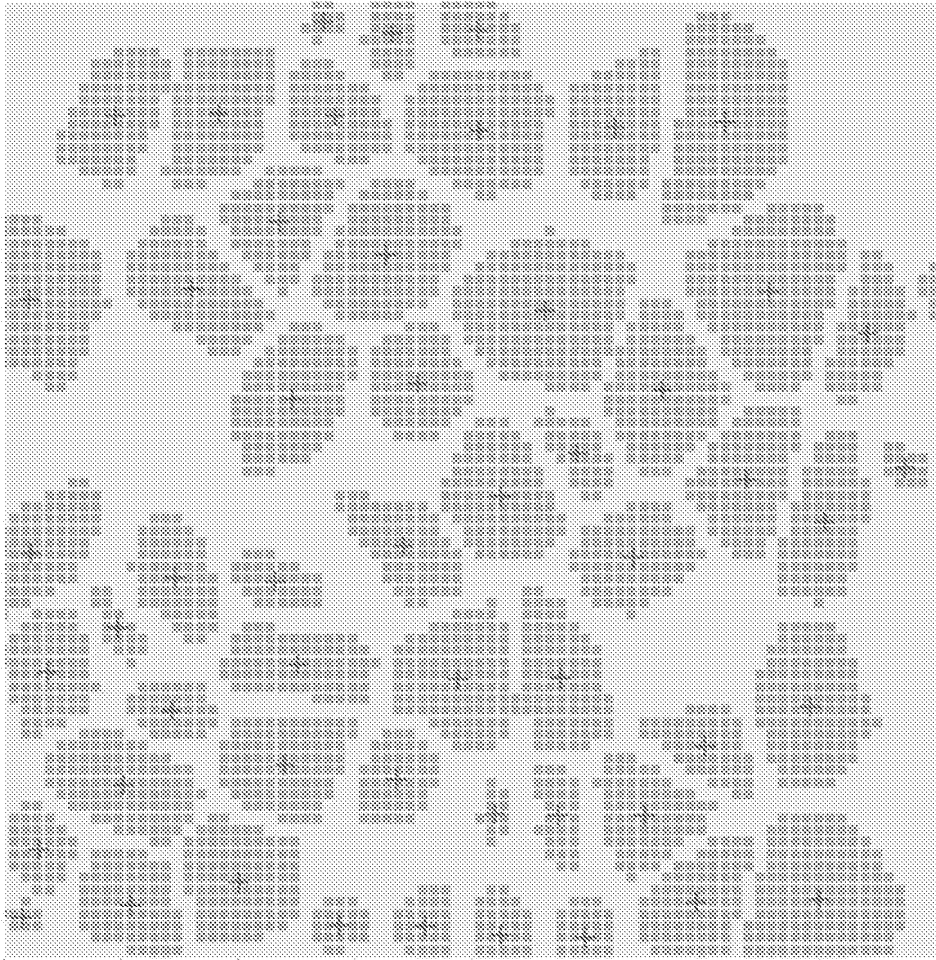


Figure 8b

Cluster Metadata 902



Spatial Distribution of Clusters – Cluster Centers, Shapes, Sizes, Background, and/or Boundaries

Figure 9

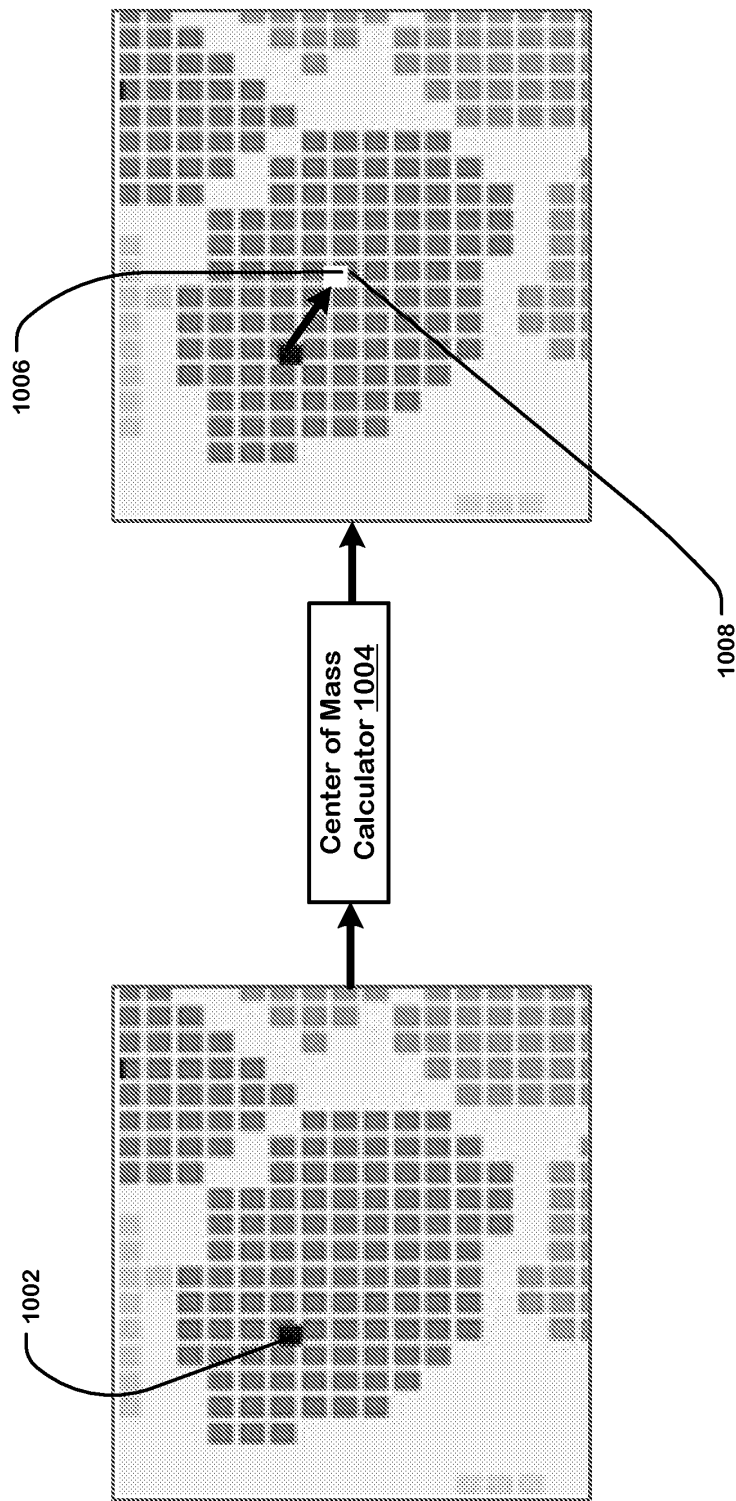


Figure 10

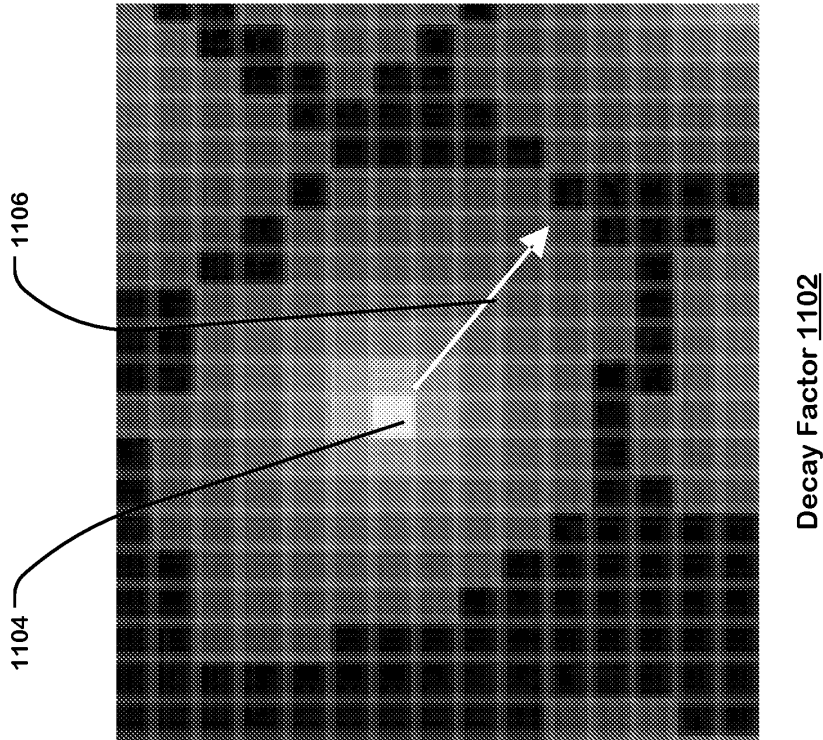


Figure 11

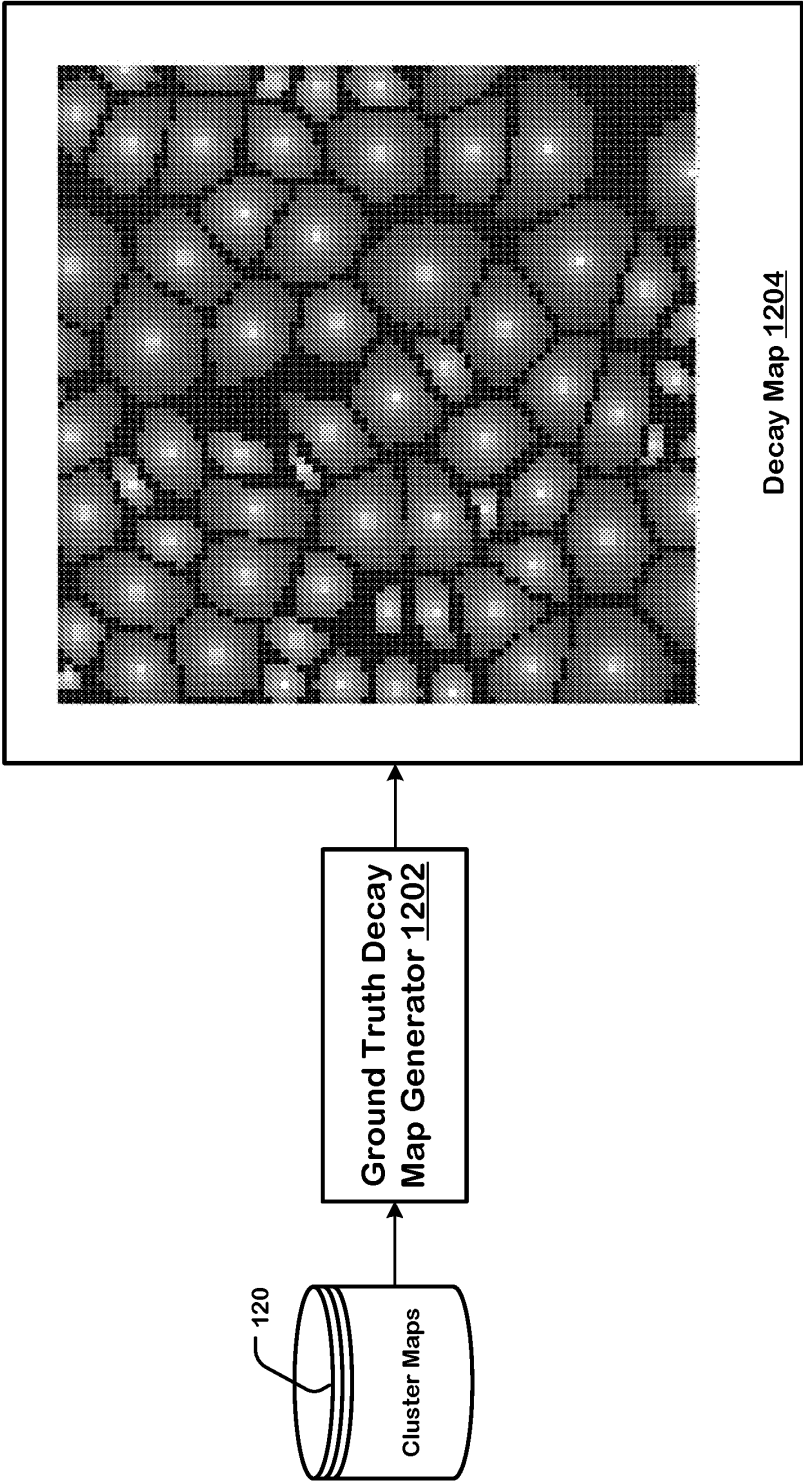


Figure 12

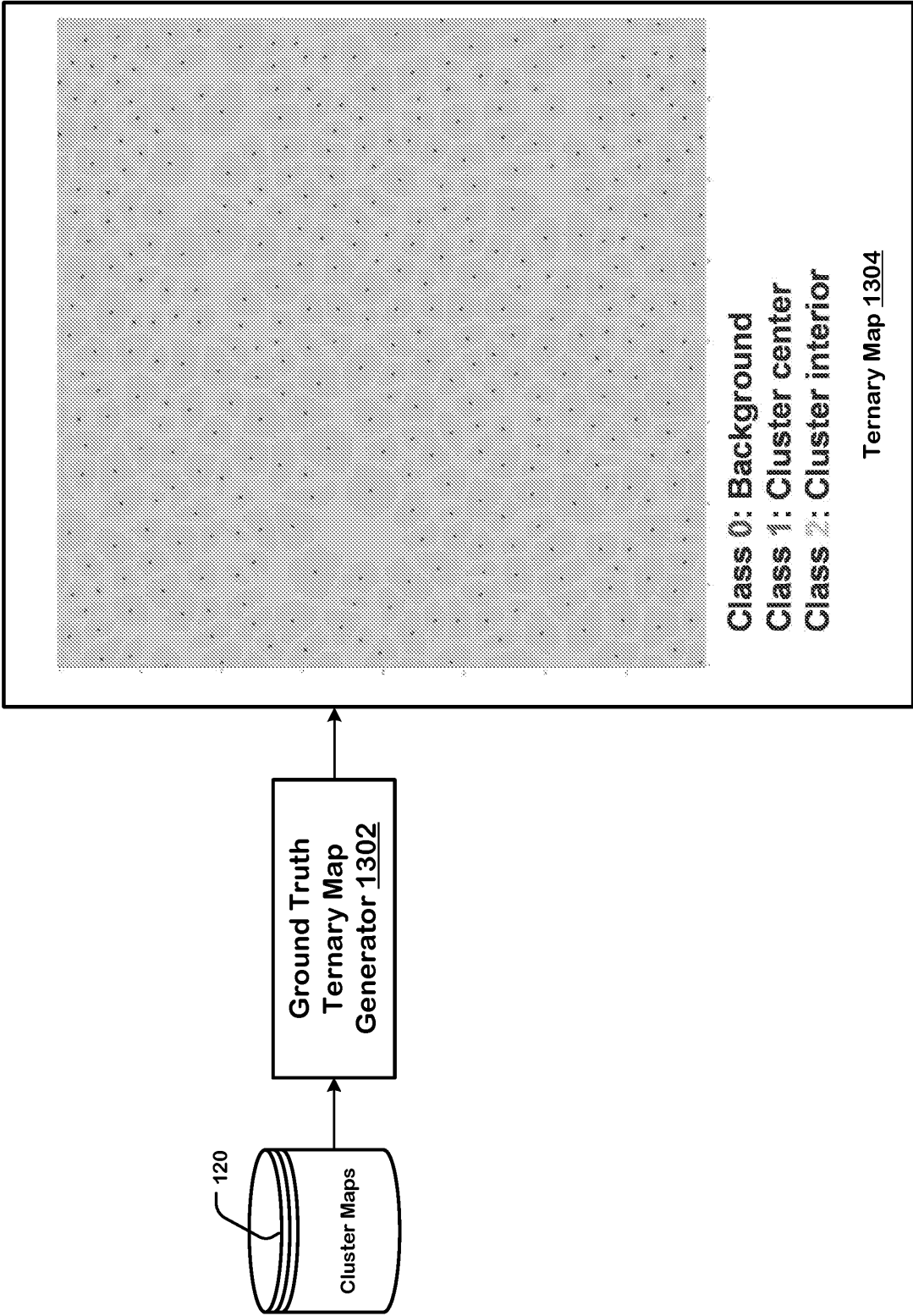


Figure 13

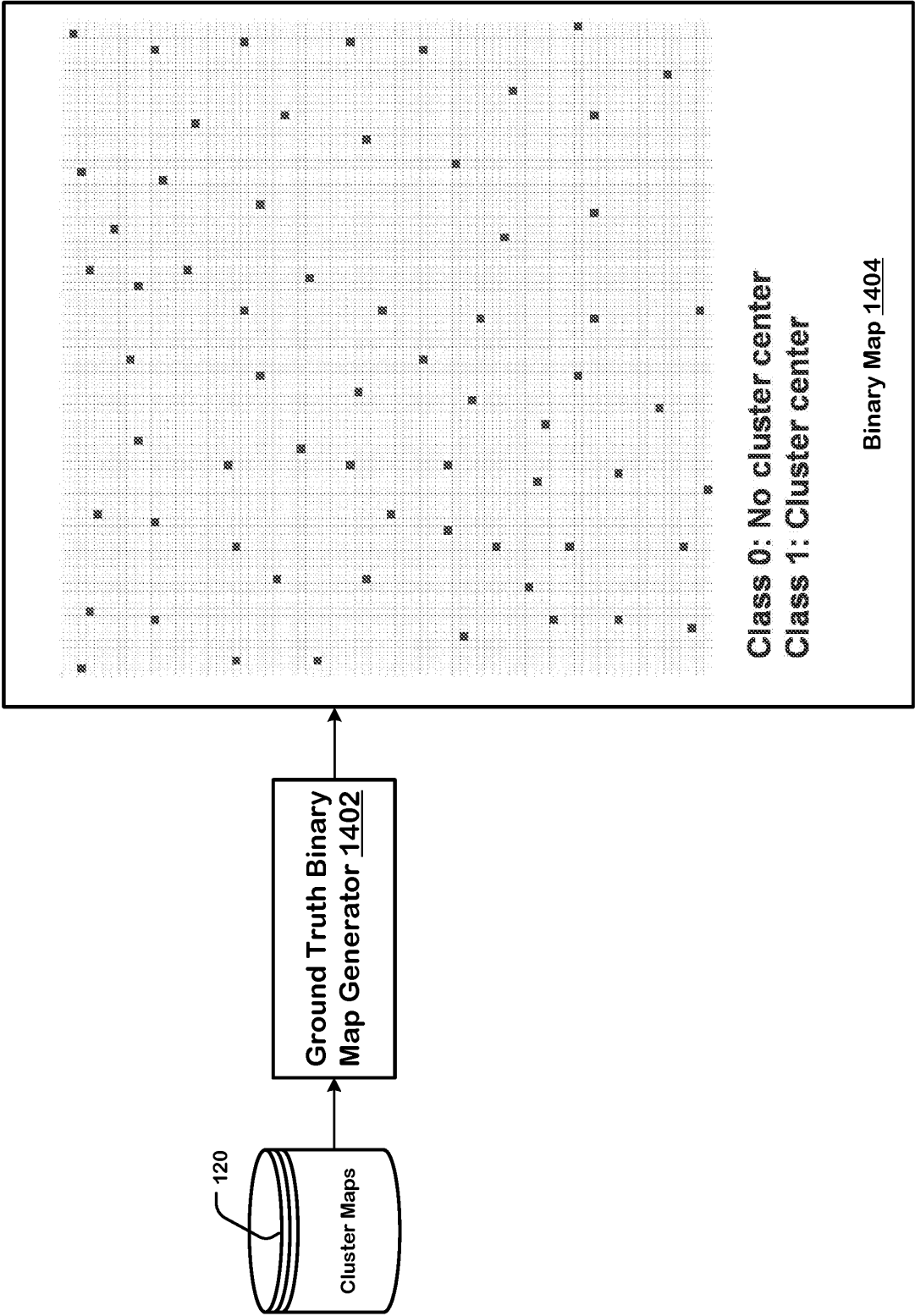


Figure 14

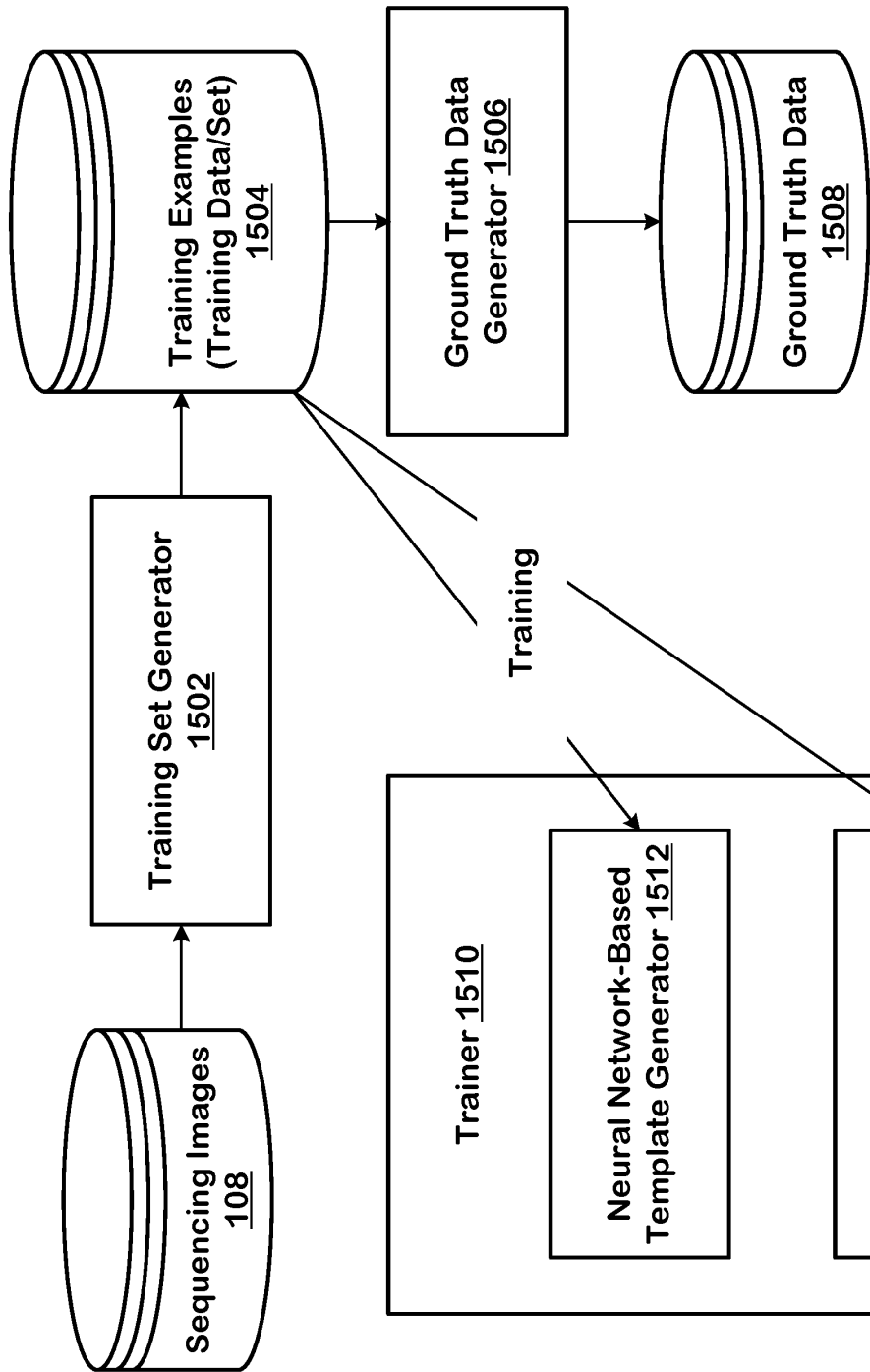


Figure 15

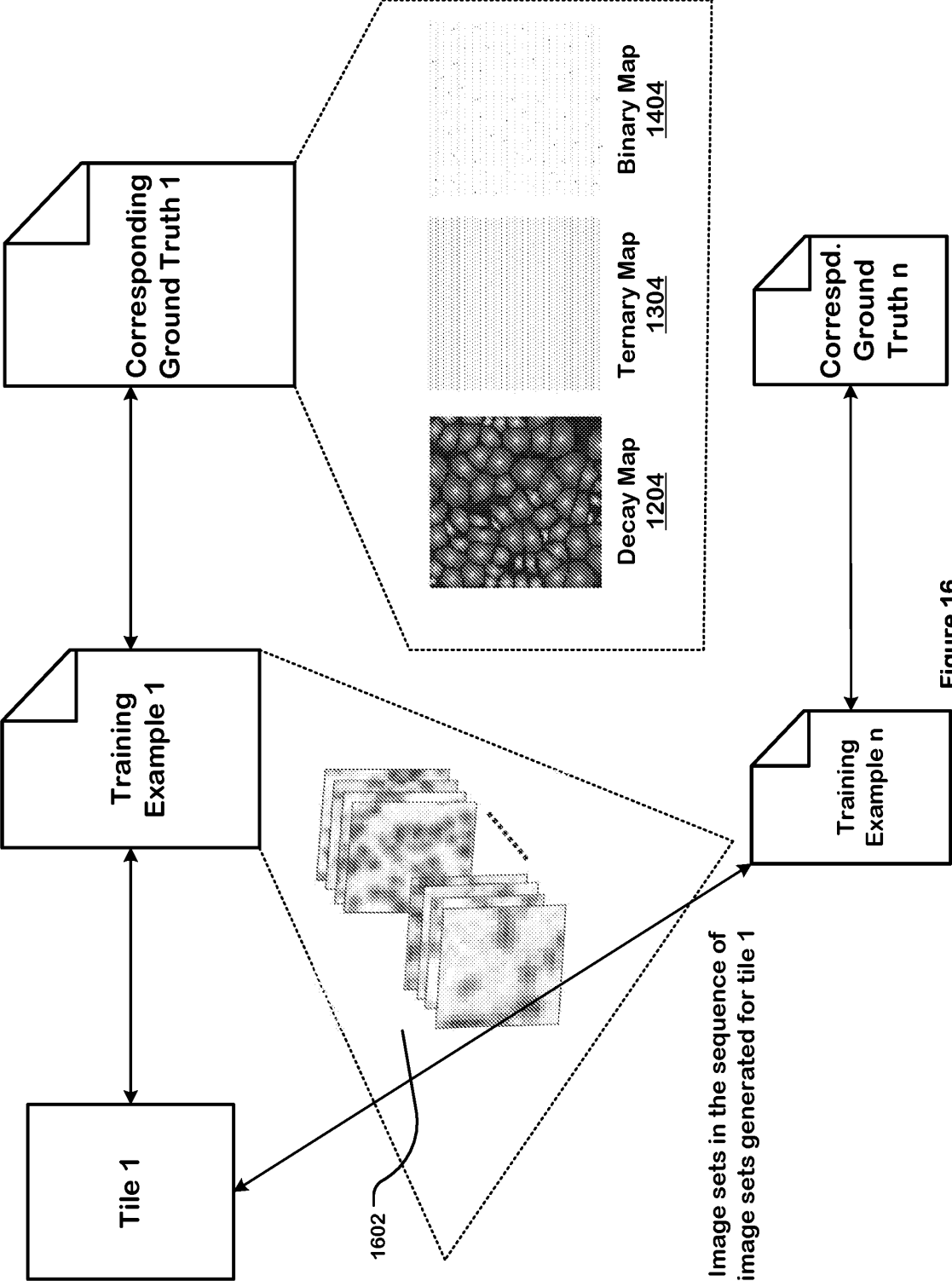


Figure 16

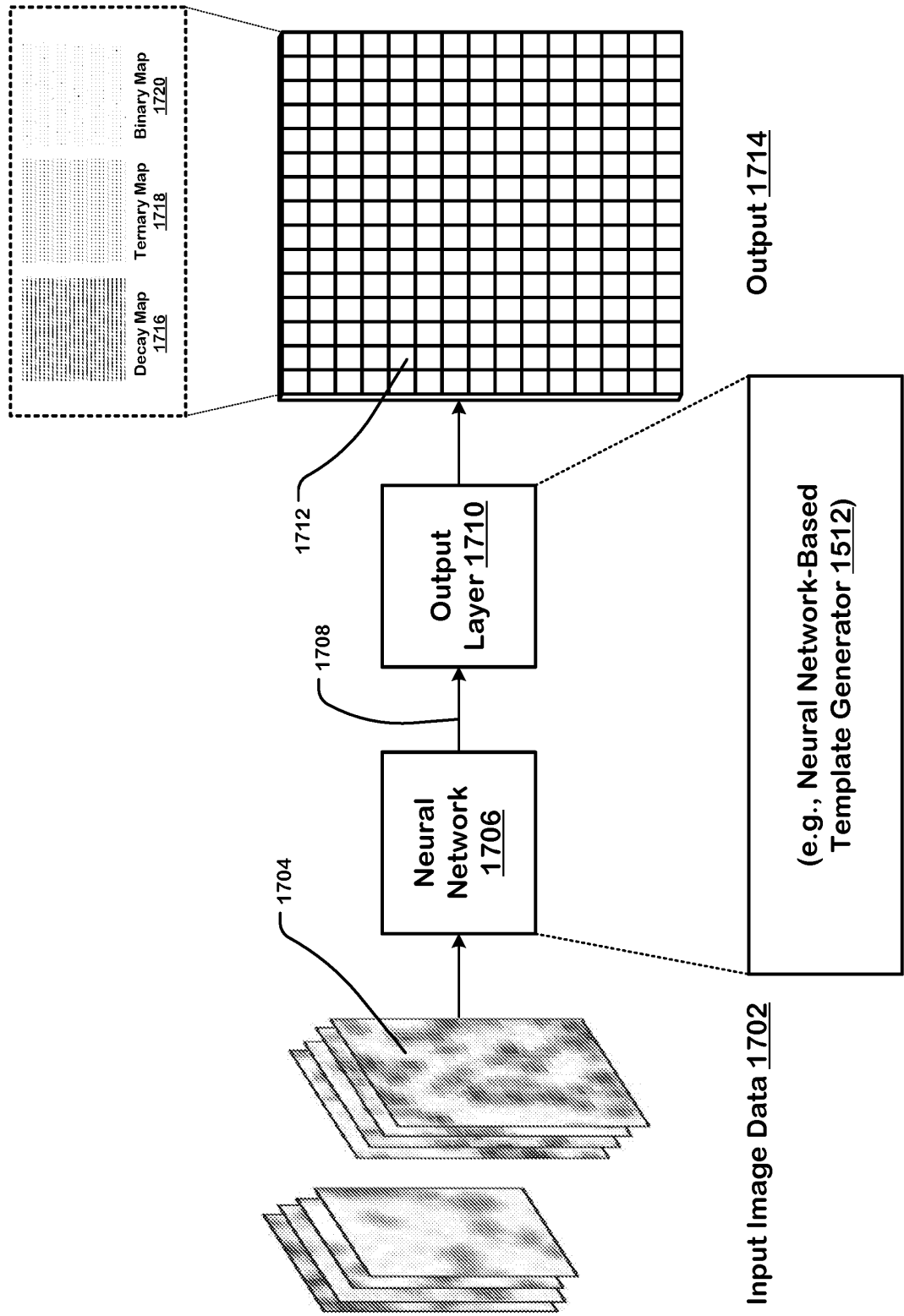


Figure 17

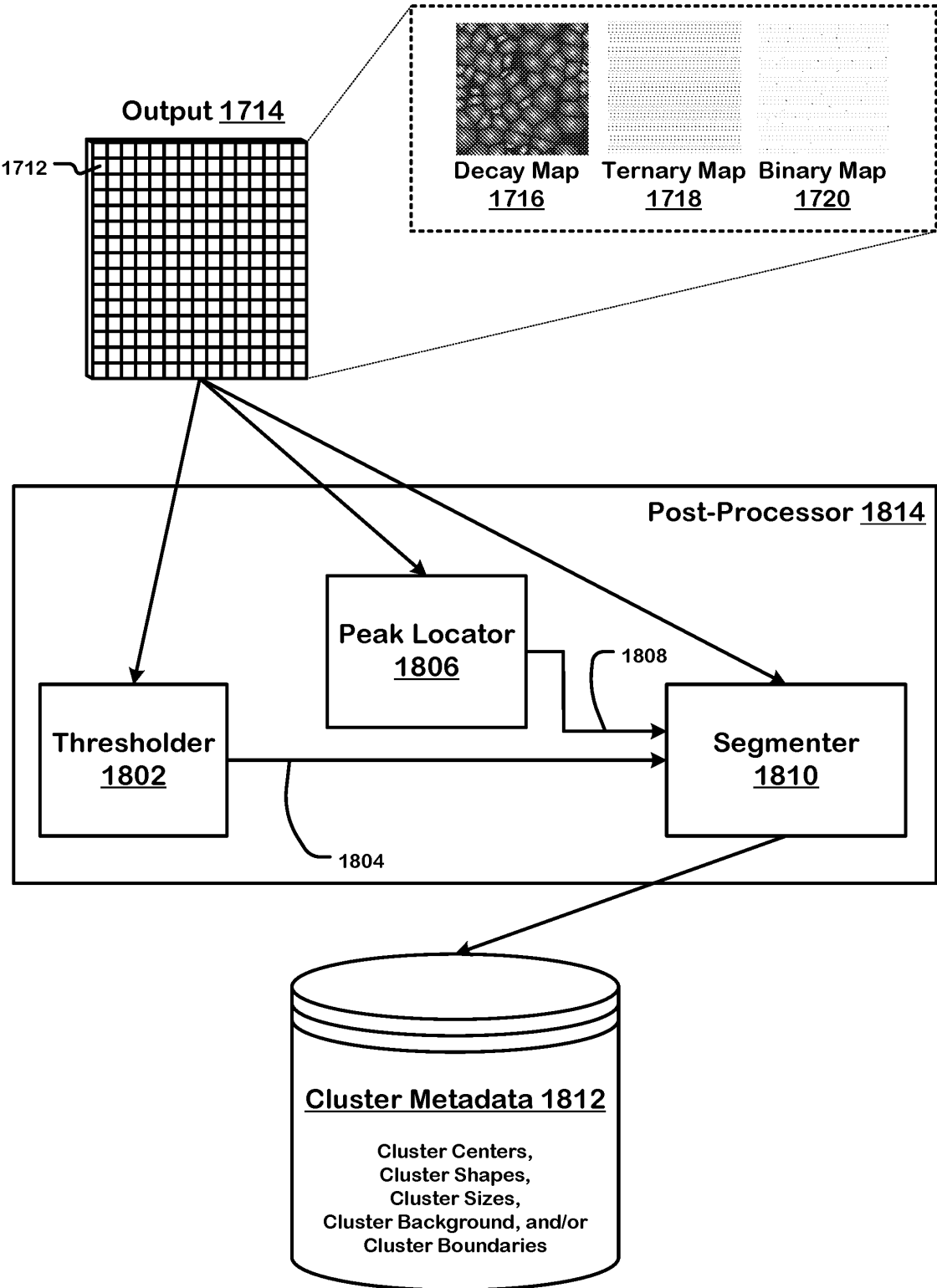


Figure 18

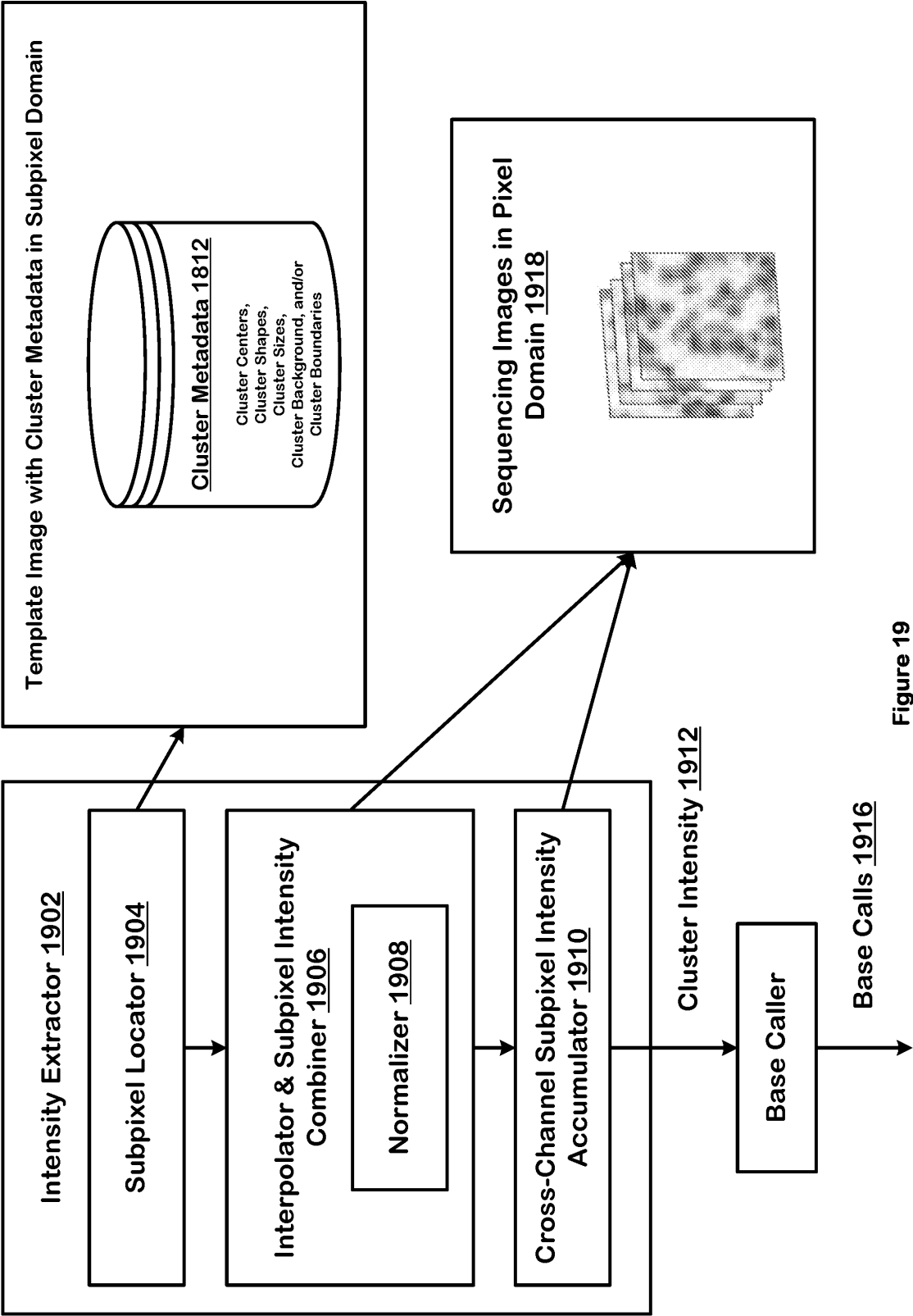


Figure 19

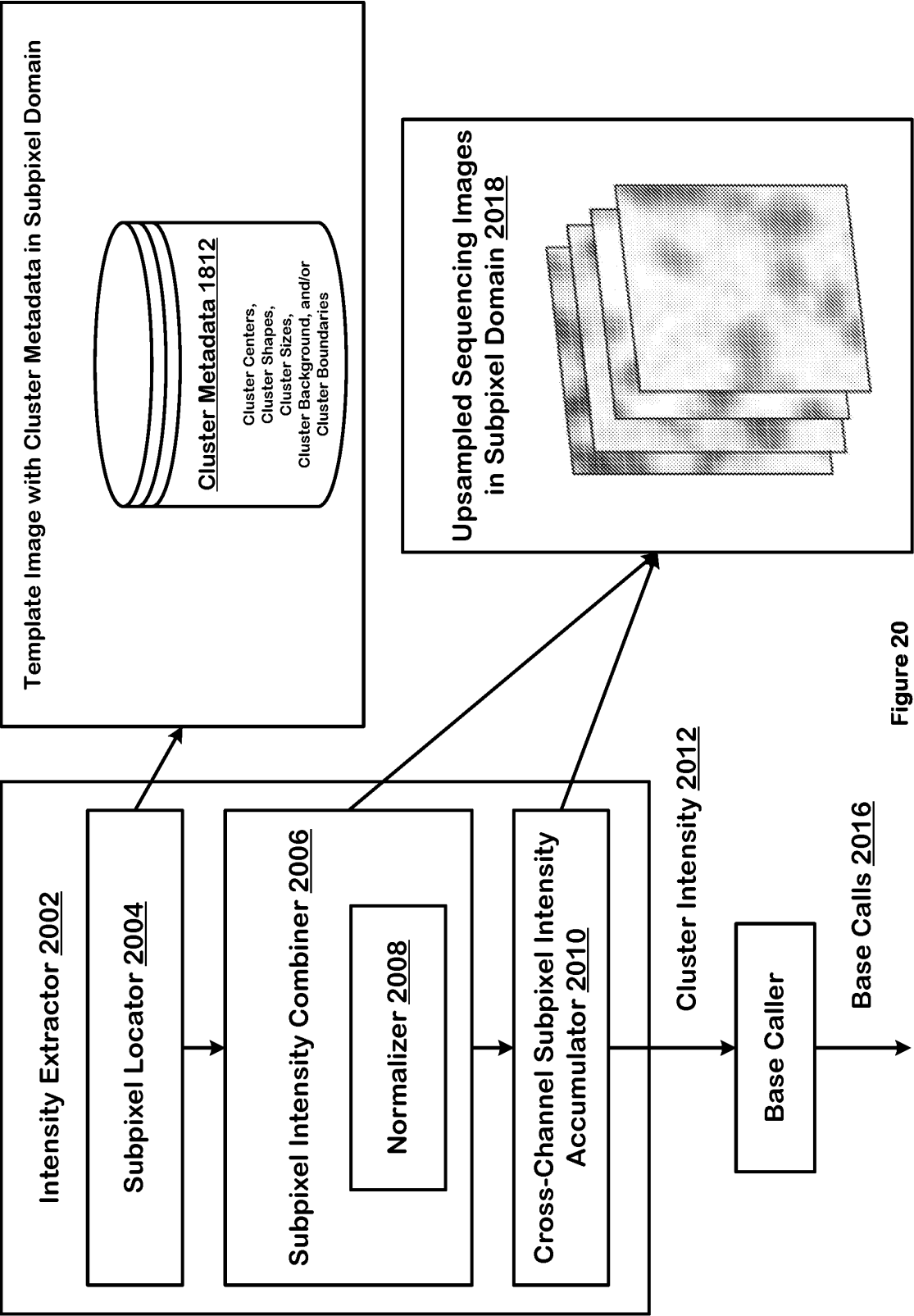


Figure 20

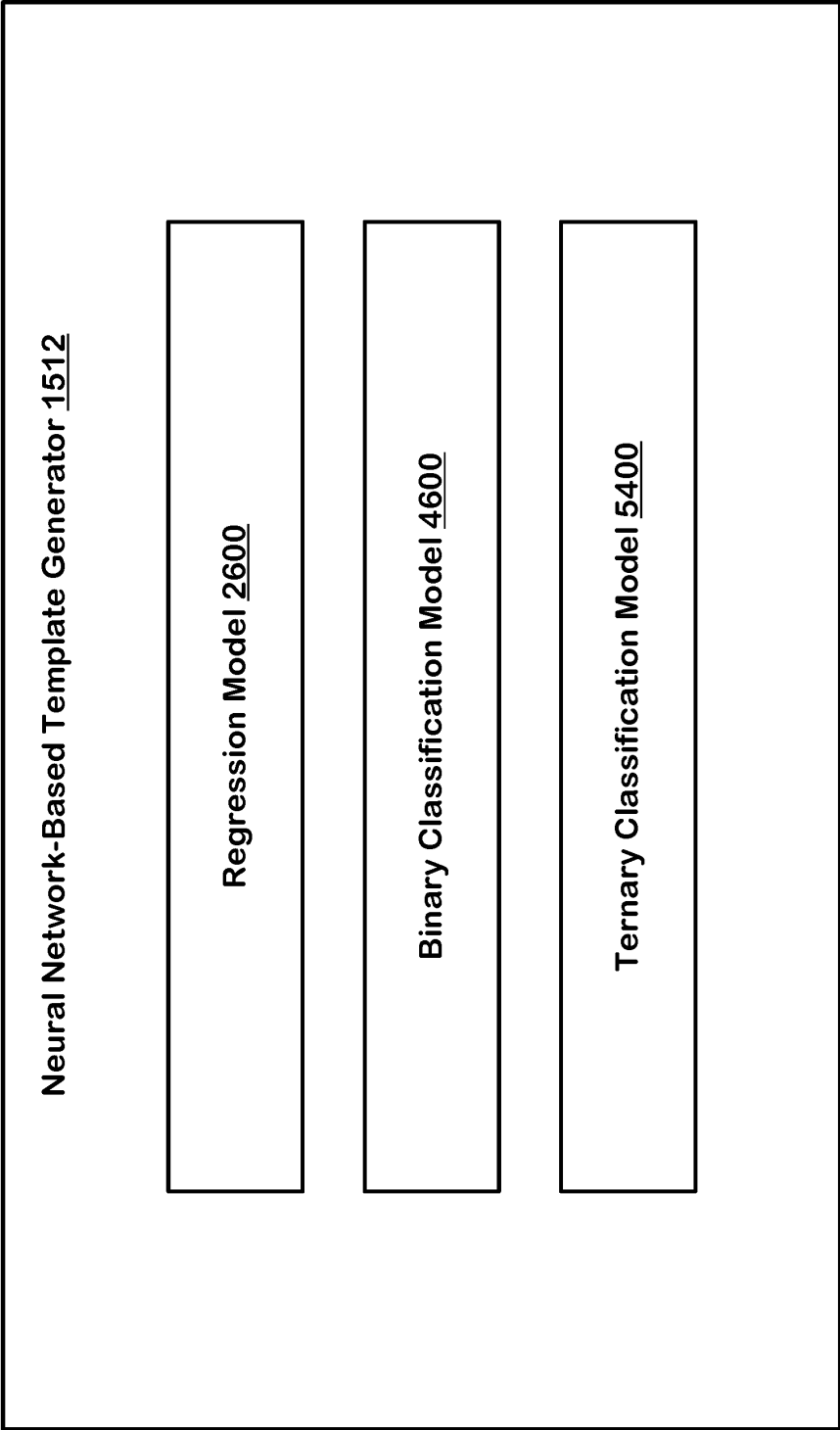
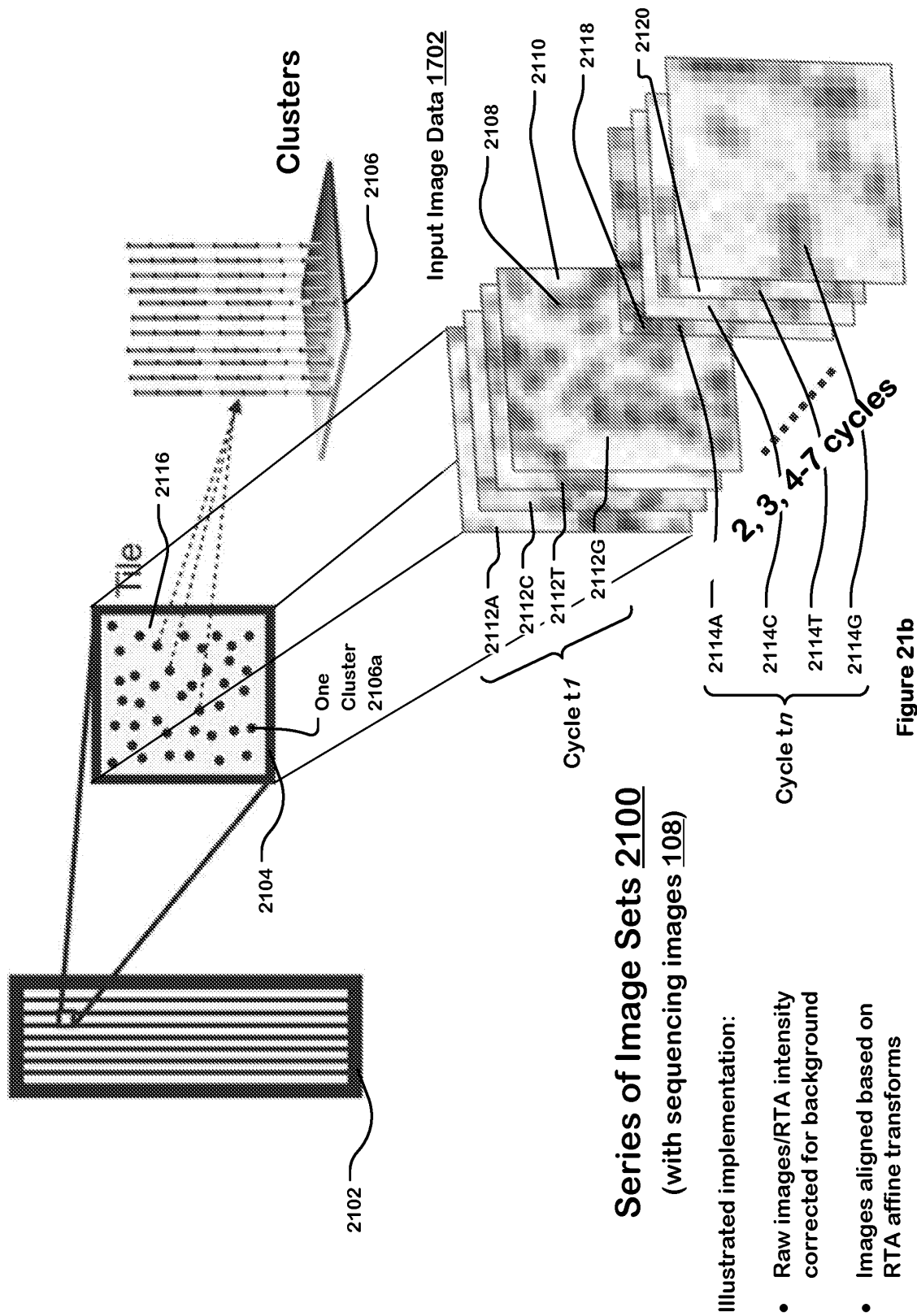


Figure 21a



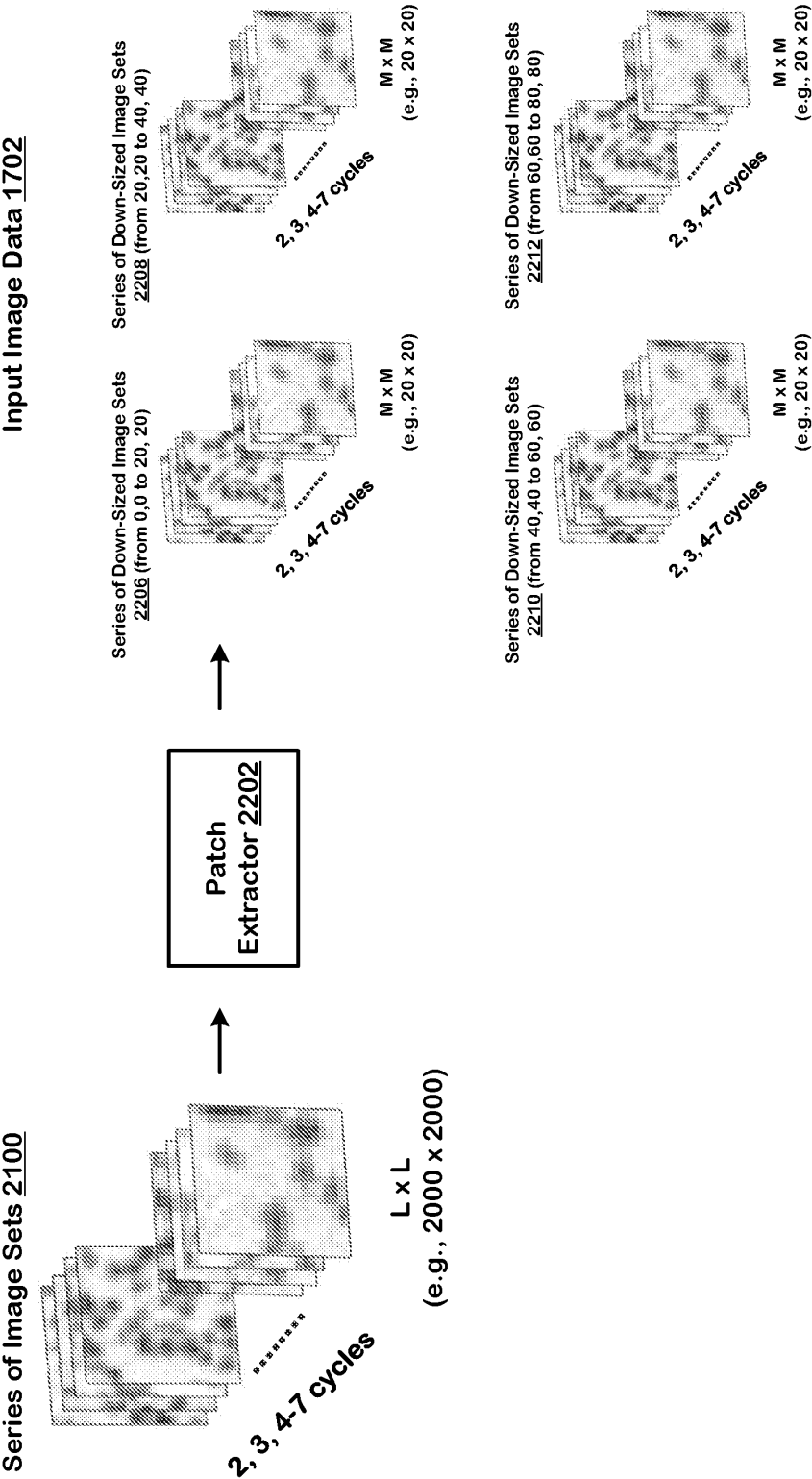


Figure 22

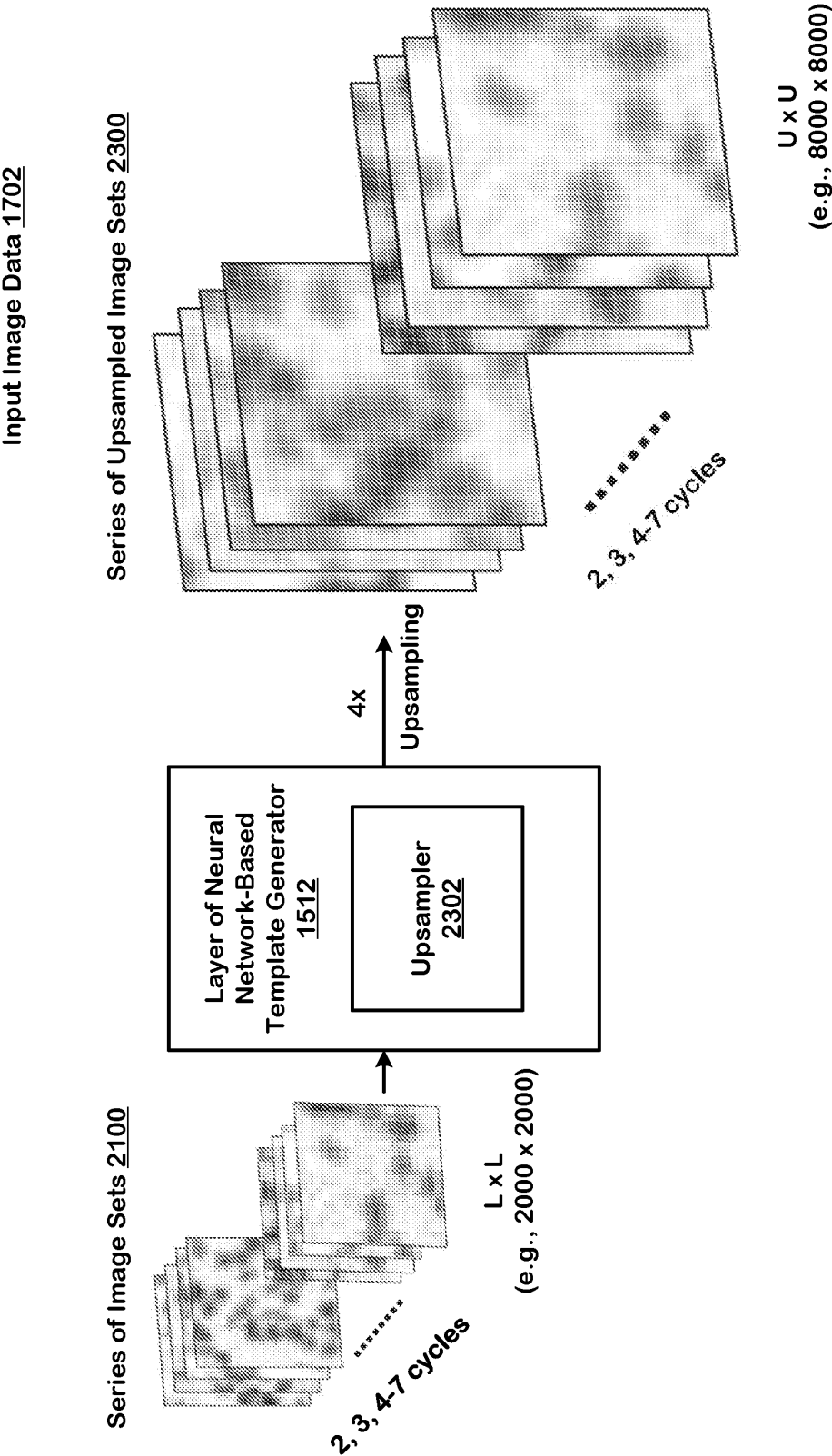


Figure 23

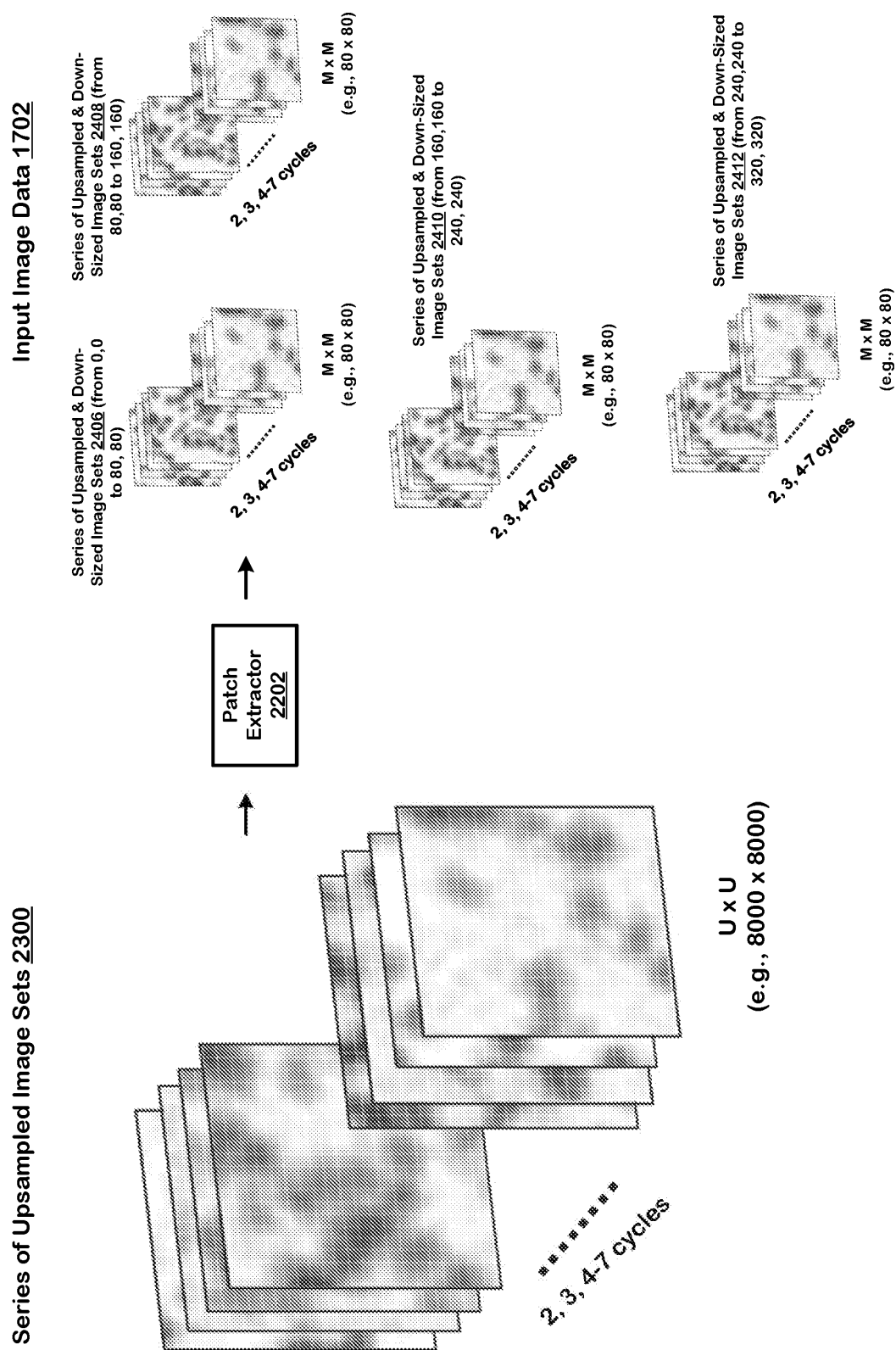
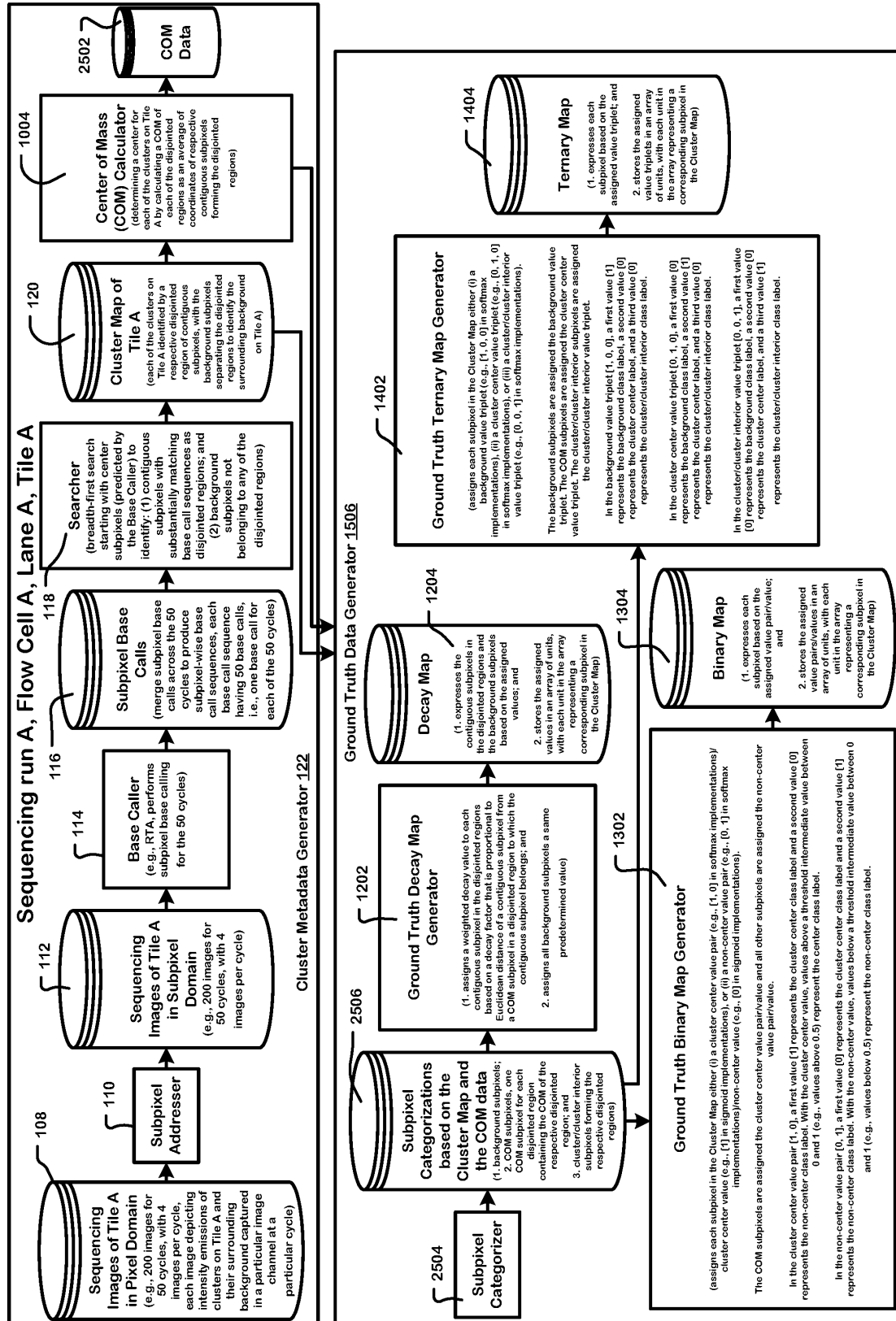
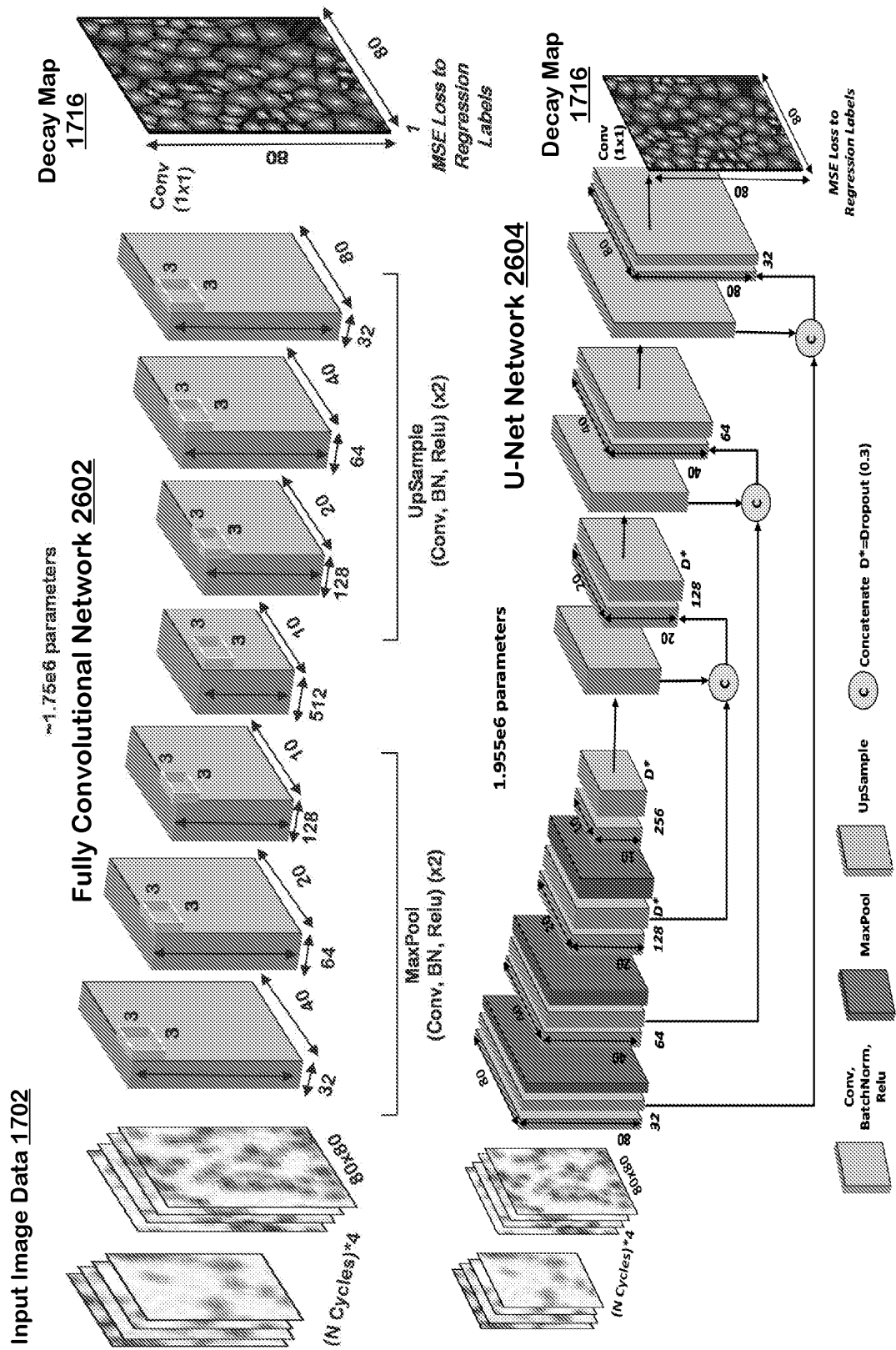


Figure 24





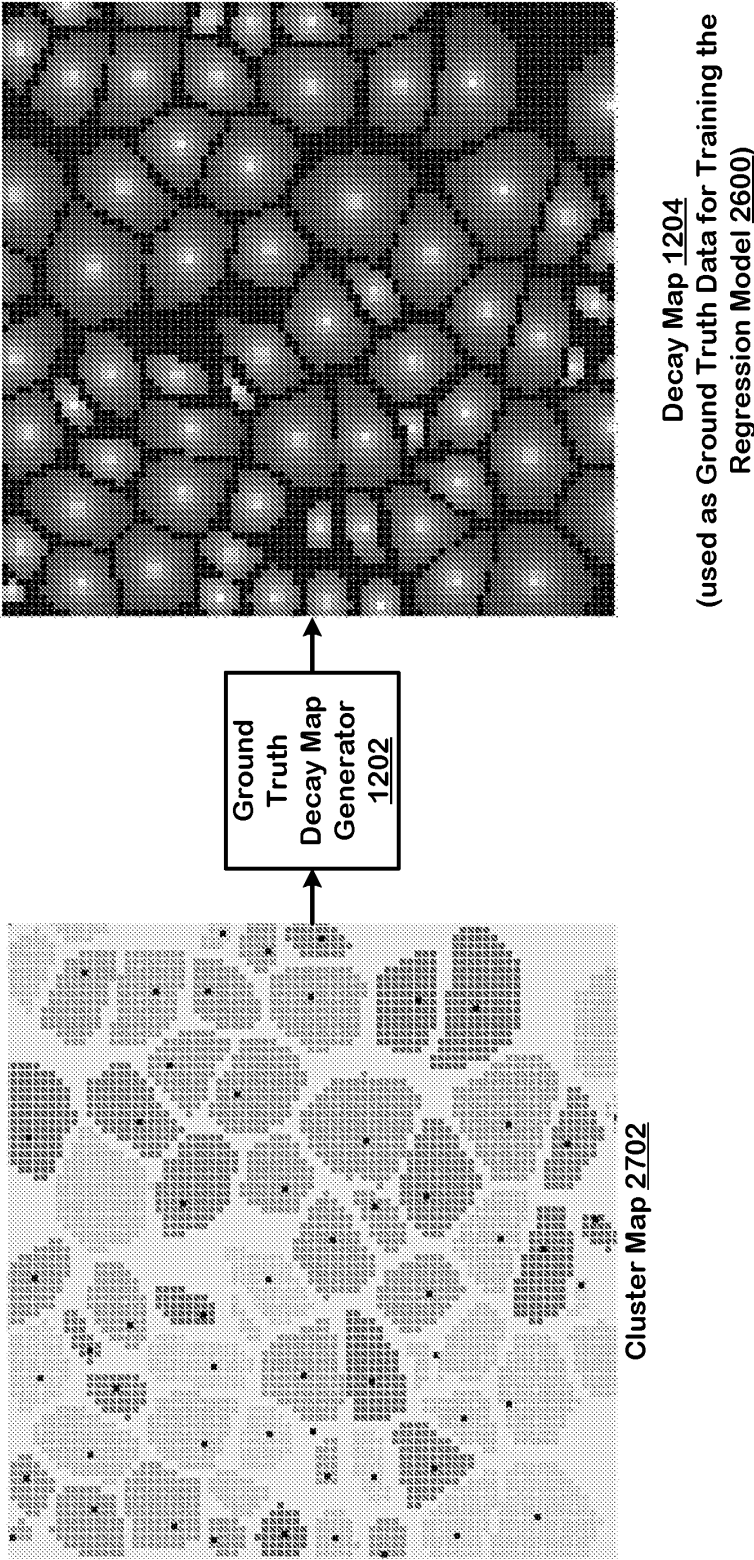


Figure 27

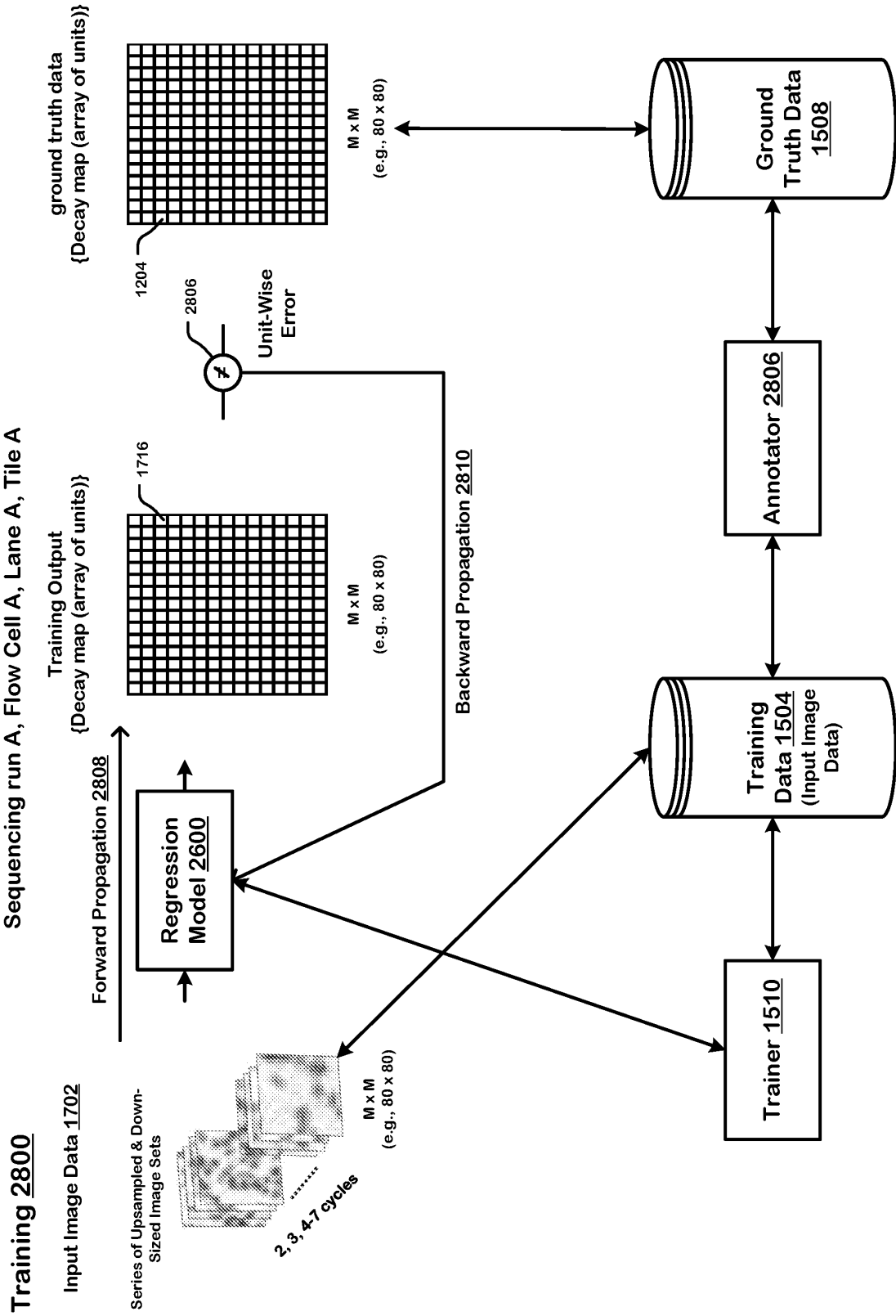


Figure 28

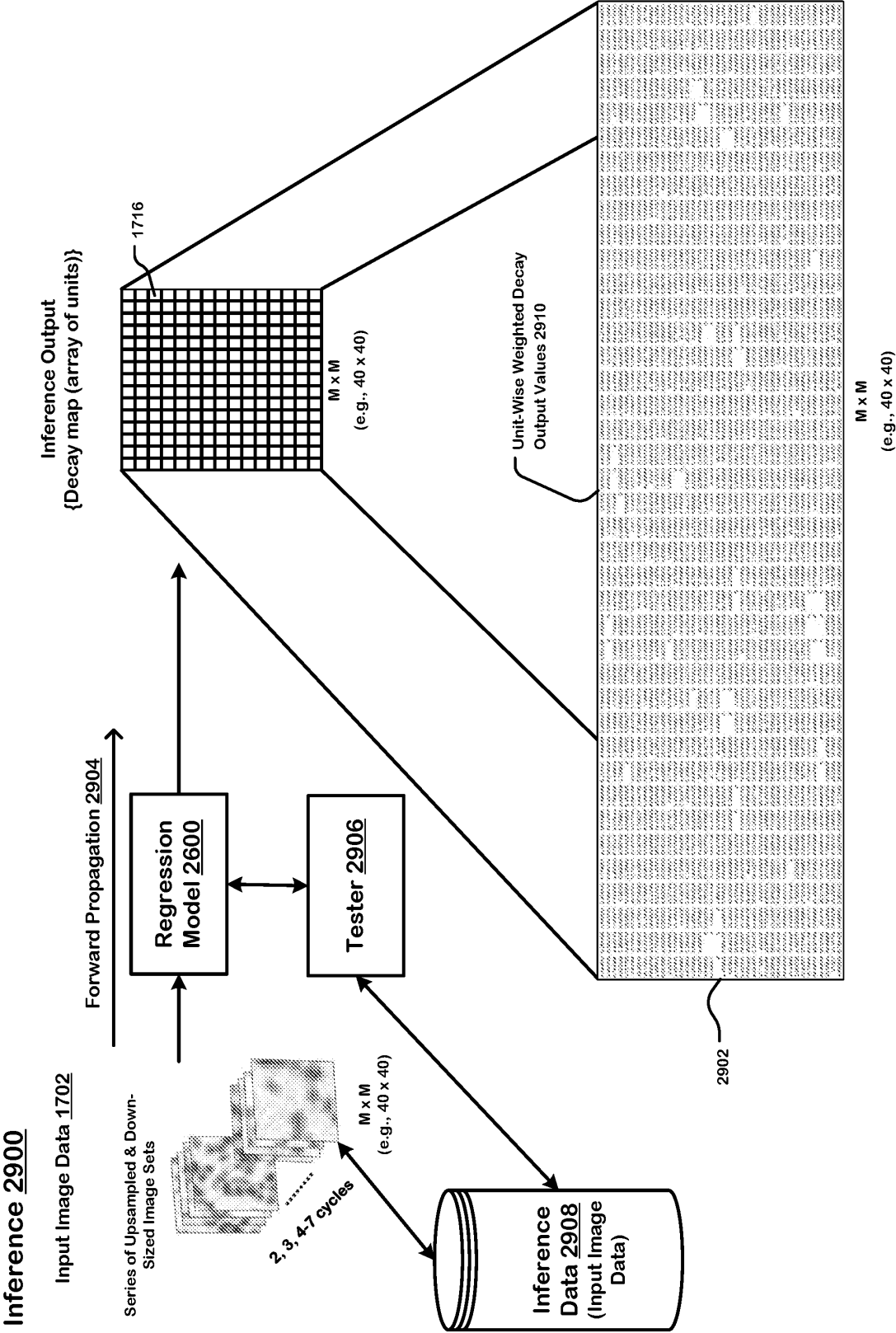


Figure 29

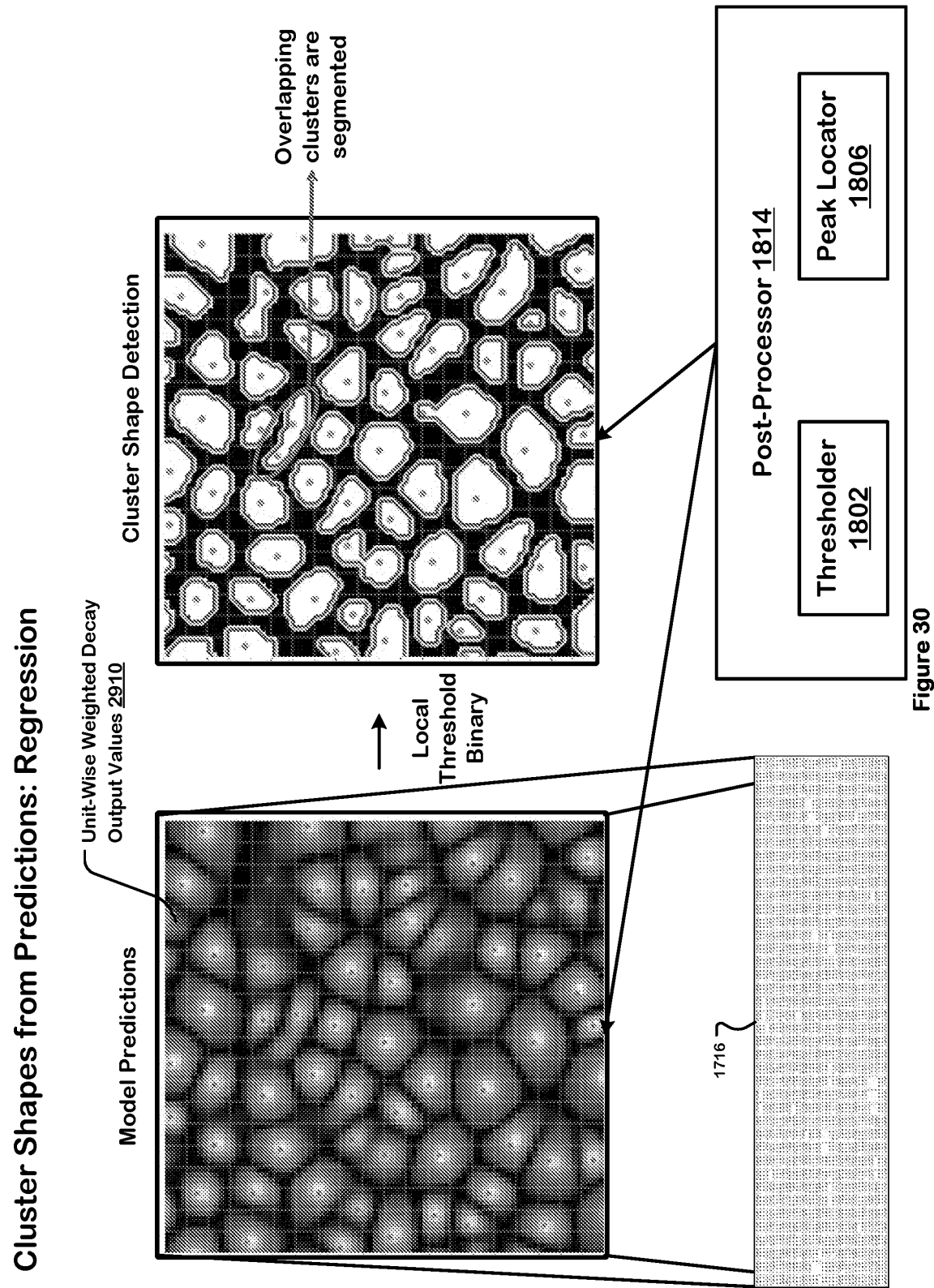


Figure 30

Watershed Segmenter for Segmentation

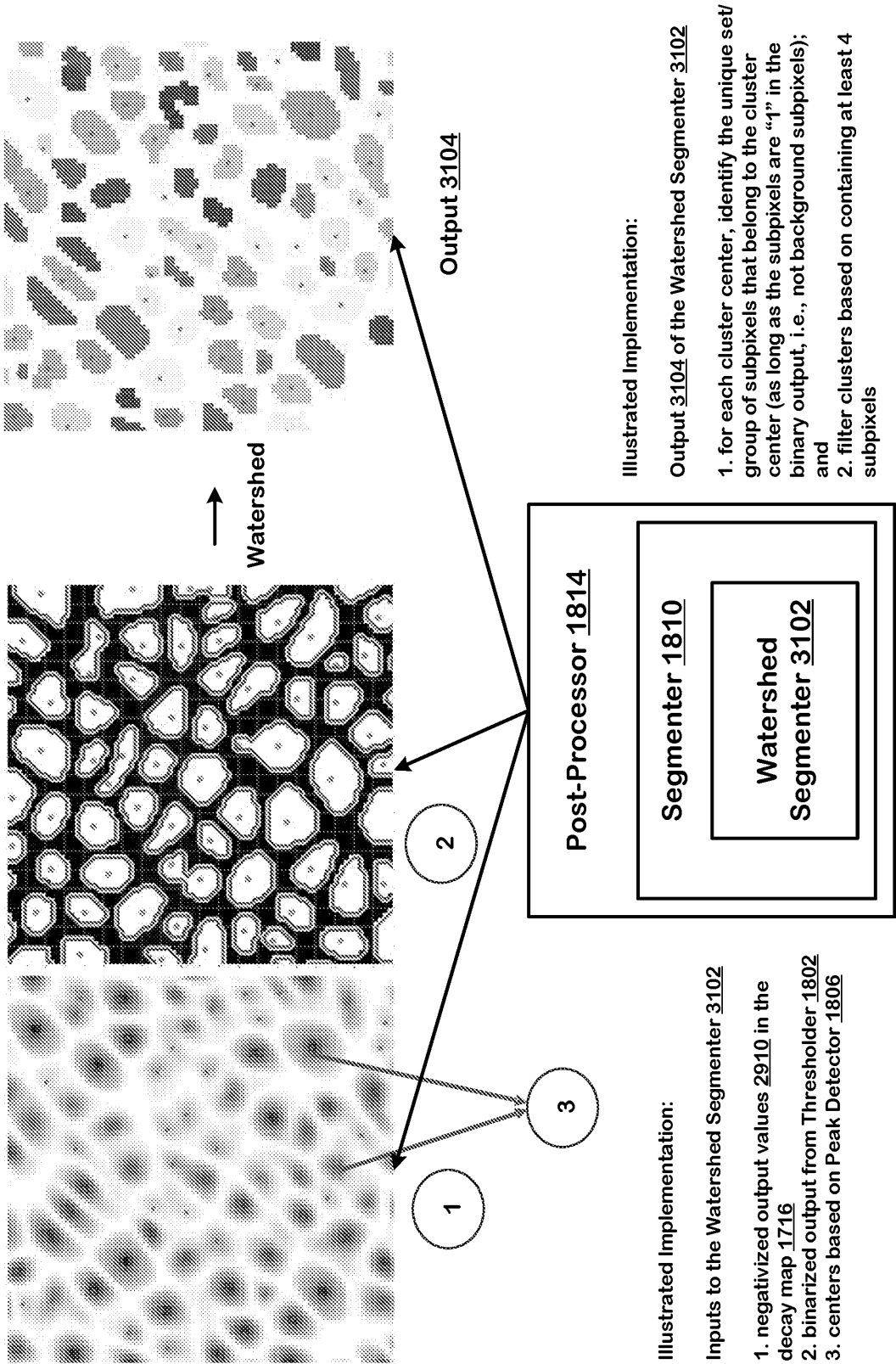
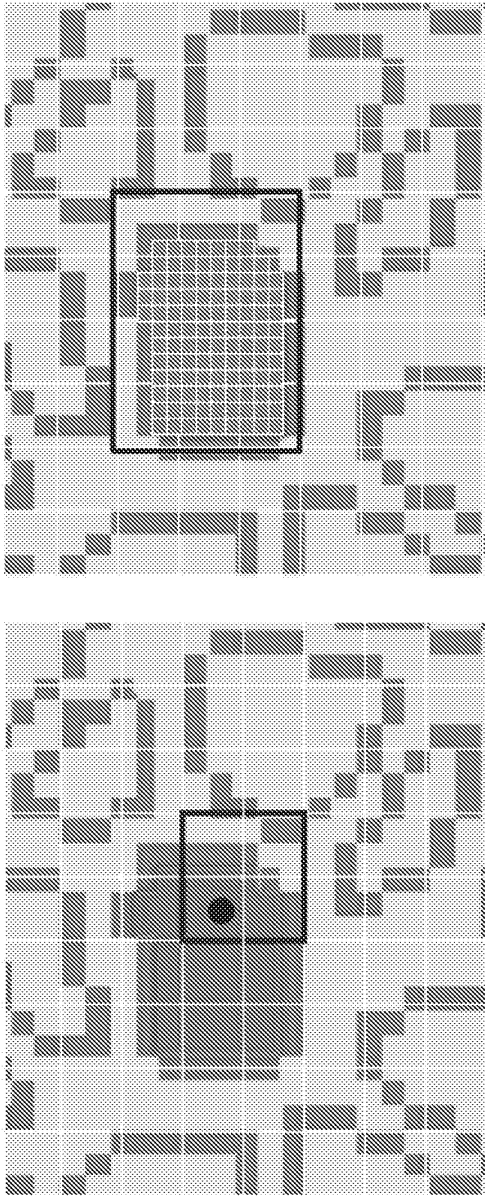


Figure 31

Layer (type)	Output Shape	Param #	Connected to
input 1 (InputLayer)	(None, None, None, 28)	0	
batch normalization 1 (BatchNormali	(None, None, None, 28)	112	input 1[0][0]
conv2d 1 (Conv2D)	(None, None, None, 32)	8994	batch normalization 1[0][0]
batch normalization 2 (BatchNormali	(None, None, None, 32)	128	conv2d 1[0][0]
conv2d 2 (Conv2D)	(None, None, None, 32)	9216	batch normalization 2[0][0]
batch normalization 3 (BatchNormali	(None, None, None, 32)	128	conv2d 2[0][0]
max pooling2d 1 (MaxPooling2D)	(None, None, None, 32)	0	batch normalization 3[0][0]
conv2d 3 (Conv2D)	(None, None, None, 64)	18432	max pooling2d 1[0][0]
batch normalization 4 (BatchNormali	(None, None, None, 64)	256	conv2d 3[0][0]
conv2d 4 (Conv2D)	(None, None, None, 64)	36864	batch normalization 4[0][0]
batch normalization 5 (BatchNormali	(None, None, None, 64)	256	conv2d 4[0][0]
max pooling2d 2 (MaxPooling2D)	(None, None, None, 64)	0	batch normalization 5[0][0]
conv2d 5 (Conv2D)	(None, None, None, 128)	73728	max pooling2d 2[0][0]
batch normalization 6 (BatchNormali	(None, None, None, 128)	512	conv2d 5[0][0]
conv2d 6 (Conv2D)	(None, None, None, 128)	147456	batch normalization 6[0][0]
batch normalization 7 (BatchNormali	(None, None, None, 128)	512	conv2d 6[0][0]
dropout 1 (Dropout)	(None, None, None, 128)	0	batch normalization 7[0][0]
max pooling2d 3 (MaxPooling2D)	(None, None, None, 128)	0	dropout 1[0][0]
conv2d 7 (Conv2D)	(None, None, None, 256)	294912	max pooling2d 3[0][0]
batch normalization 8 (BatchNormali	(None, None, None, 256)	1024	conv2d 7[0][0]
conv2d 8 (Conv2D)	(None, None, None, 256)	655328	batch normalization 8[0][0]
batch normalization 9 (BatchNormali	(None, None, None, 256)	1024	conv2d 8[0][0]
dropout 2 (Dropout)	(None, None, None, 256)	0	batch normalization 9[0][0]
up sampling2d 1 (UpSampling2D)	(None, None, None, 256)	0	dropout 2[0][0]
concatenate 1 (Concatenate)	(None, None, None, 384)	0	up sampling2d 1[0][0] batch normalization 7[0][0]
conv2d 9 (Conv2D)	(None, None, None, 128)	442368	concatenate 1[0][0]
batch normalization 10 (BatchNormal	(None, None, None, 128)	512	conv2d 9[0][0]
conv2d 10 (Conv2D)	(None, None, None, 128)	147456	batch normalization 10[0][0]
batch normalization 11 (BatchNormal	(None, None, None, 128)	512	conv2d 10[0][0]
dropout 3 (Dropout)	(None, None, None, 128)	0	batch normalization 11[0][0]
up sampling2d 2 (UpSampling2D)	(None, None, None, 128)	0	dropout 3[0][0]
concatenate 2 (Concatenate)	(None, None, None, 192)	0	up sampling2d 2[0][0] batch normalization 9[0][0]
conv2d 11 (Conv2D)	(None, None, None, 64)	110592	concatenate 2[0][0]
batch normalization 12 (BatchNormal	(None, None, None, 64)	256	conv2d 11[0][0]
conv2d 12 (Conv2D)	(None, None, None, 64)	26864	batch normalization 12[0][0]
batch normalization 13 (BatchNormal	(None, None, None, 64)	256	conv2d 12[0][0]
up sampling2d 3 (UpSampling2D)	(None, None, None, 64)	0	batch normalization 13[0][0]
concatenate 3 (Concatenate)	(None, None, None, 96)	0	up sampling2d 3[0][0] batch normalization 9[0][0]
conv2d 13 (Conv2D)	(None, None, None, 32)	27648	concatenate 3[0][0]
batch normalization 14 (BatchNormal	(None, None, None, 32)	128	conv2d 13[0][0]
conv2d 14 (Conv2D)	(None, None, None, 32)	8216	batch normalization 14[0][0]
batch normalization 15 (BatchNormal	(None, None, None, 32)	128	conv2d 14[0][0]
pred (Conv2D)	(None, None, None, 1)	32	batch normalization 15[0][0]
Total params: 1,955,416			
Trainable params: 1,955,404			
Non-trainable params: 12			

Figure 32

Intensity Extraction using Cluster Shapes



Approach 1	$\frac{\sum_{\text{subpixels}} \text{Bilinear Interpolated Intensity from Original Image}}{\text{No. of Subpixels in Cluster}}$
Approach 2	$\frac{\sum_{\text{pixels}} \text{intensity} * \frac{\text{no of subpixels}}{16}}{\sum_{\text{pixels}} \frac{\text{no of subpixels}}{16}}$
Approach 3	$\frac{\sum_{\text{pixels}} * \text{Intensity from Upsampled Image}}{\text{No. of pixels in Cluster}}$

* These pixels are in upsampled space. Upsampled image obtained by bicubic interpolation.

Figure 33

How good are the predicted clusters?

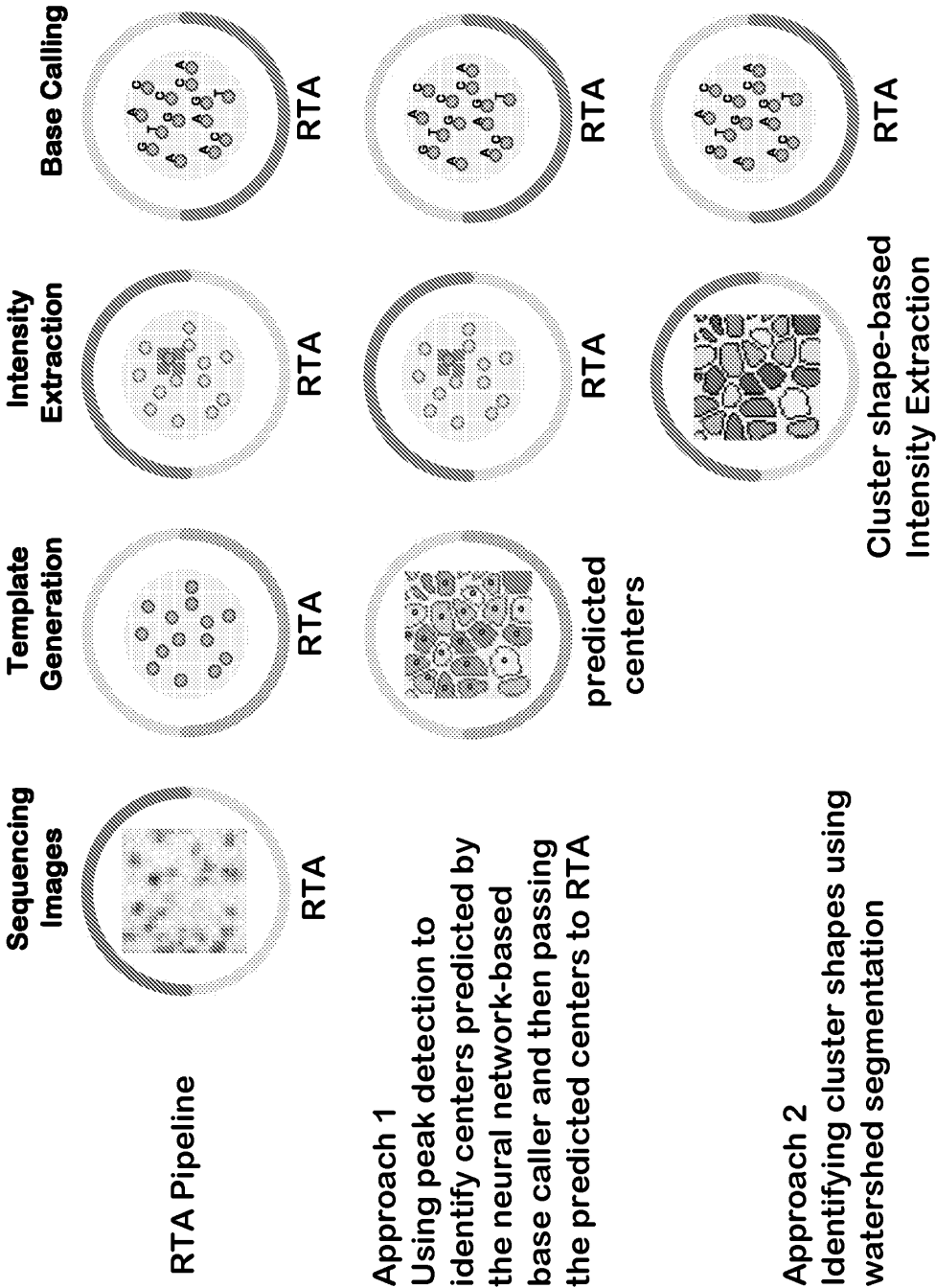


Figure 34

Centers of Mass vs. RTA centers

Centers of Mass

956048 + 0 in total (QC-passed reads + QC-failed reads)
158 + 0 secondary
0 + 0 supplementary
841 + 0 duplicates
936609 + 0 mapped (97.97% : N/A)
955890 + 0 paired in sequencing
477945 + 0 read1
477945 + 0 read2
928926 + 0 properly paired (97.18% : N/A)
930732 + 0 with itself and mate mapped
5719 + 0 singletons (0.60% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

bases mapped: 141404101 # ignores clipping
bases mapped (cigar): 140673703
bases trimmed: 0
bases duplicated: 126991
mismatches: 498799 # from NM fields
error rate: 3.545787e-03 # mismatches / bases mapped (cigar)

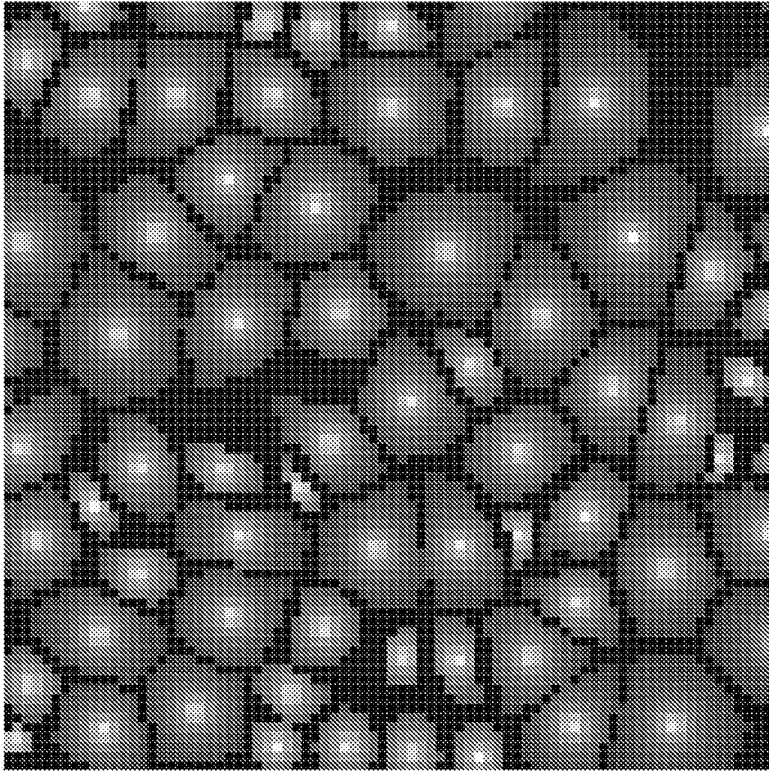
RTA centers

956054 + 0 in total (QC-passed reads + QC-failed reads)
164 + 0 secondary
0 + 0 supplementary
1045 + 0 duplicates
932910 + 0 mapped (97.58% : N/A)
955890 + 0 paired in sequencing
477945 + 0 read1
477945 + 0 read2
923322 + 0 properly paired (96.59% : N/A)
925276 + 0 with itself and mate mapped
7470 + 0 singletons (0.78% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

bases mapped: 140844646 # ignores clipping
bases mapped (cigar): 139583478
bases trimmed: 0
bases duplicated: 157795
mismatches: 821045 # from NM fields
error rate: 5.877897e-03 # mismatches / bases mapped (cigar)

Figure 35

GROUND TRUTH



MODEL PREDICTIONS

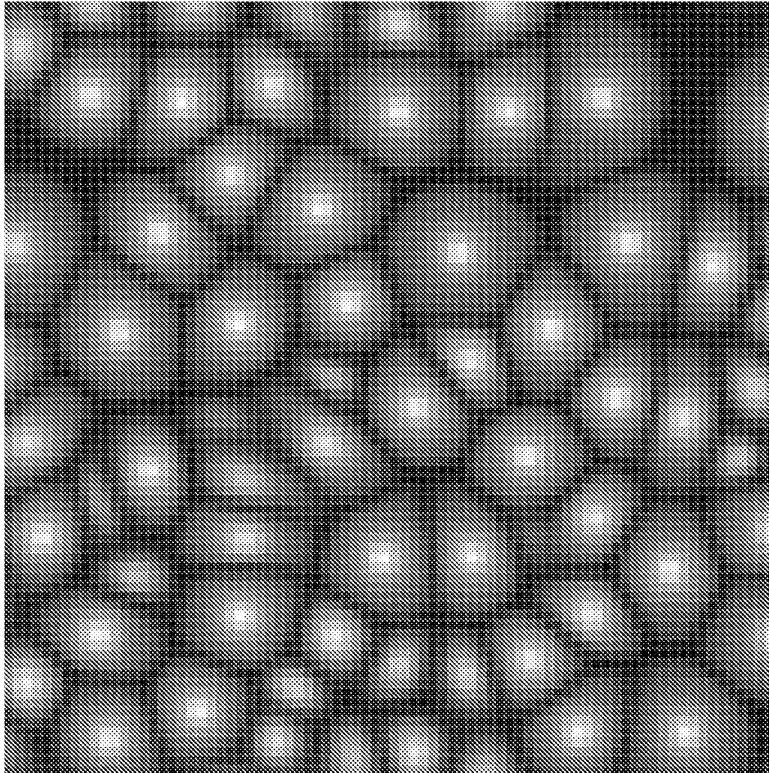
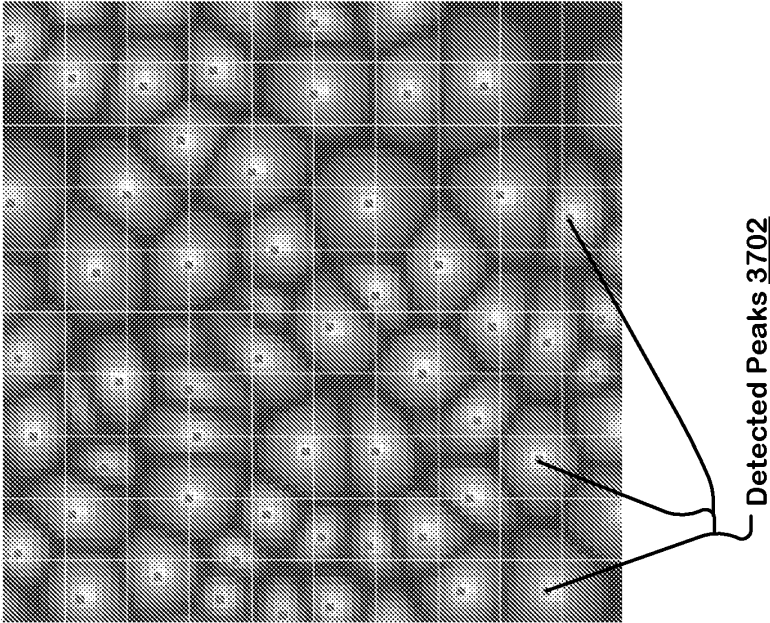


Figure 36

PEAK DETECTION



Peak Locator
1806

MODEL PREDICTIONS

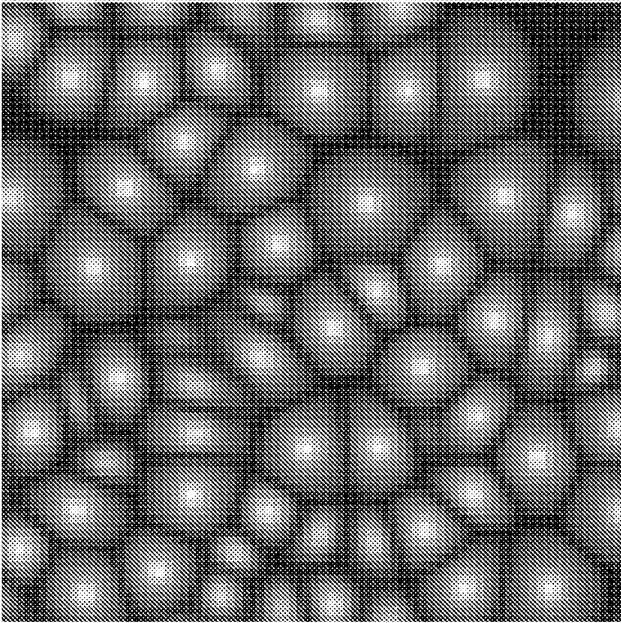
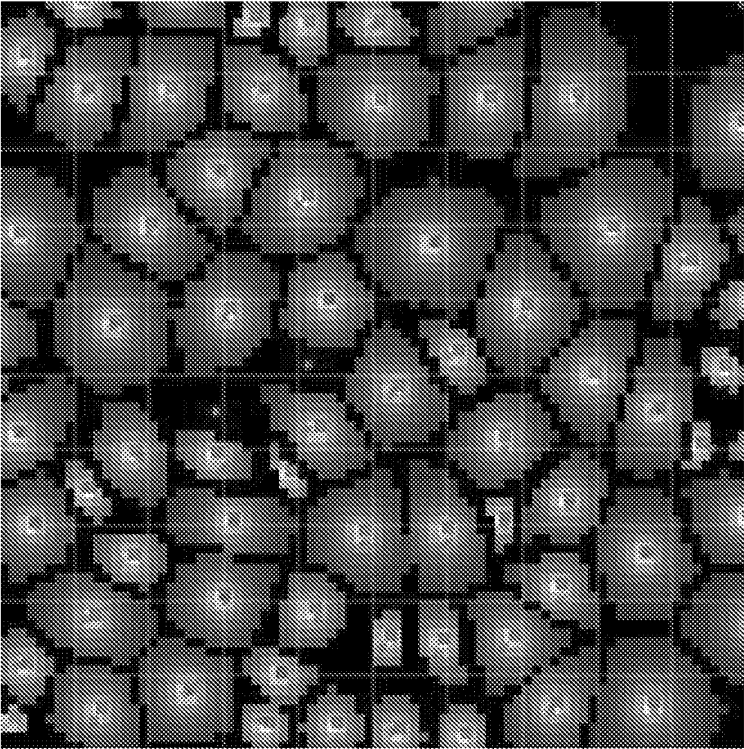


Figure 37

GROUND TRUTH



MODEL PREDICTIONS

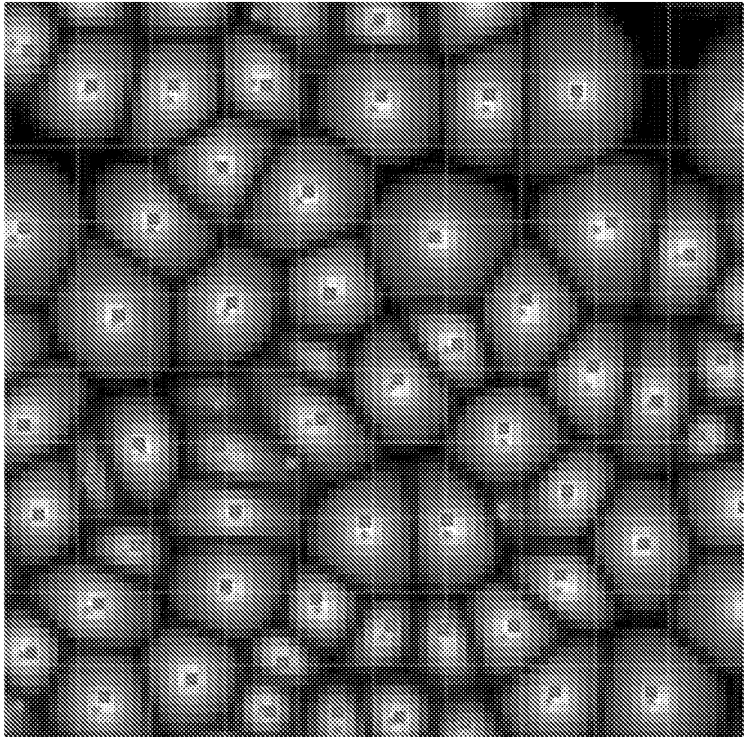
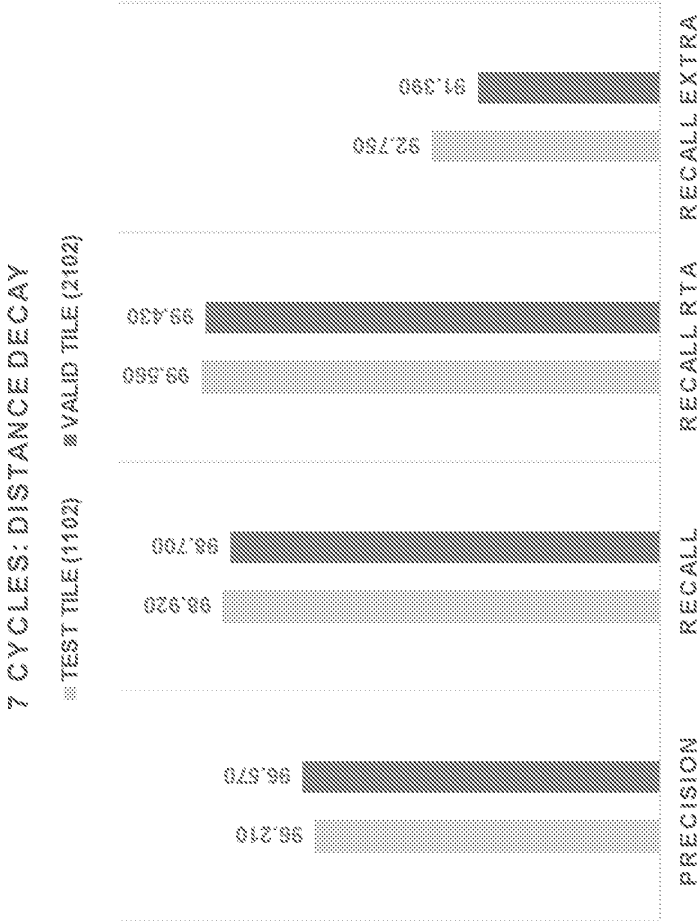


Figure 38

Regression Model: After Optimization



□ Good at recovering all identified centers

Figure 39

NORMAL RUN 20pM Tile: 1102 (7 cycles)

	RTA	Within Boundary +/- 10 (Unet)
# clusters	903,040	988,884
% PF	88.55% (799,684)	84.54% (836,197)
% Aligned	97.50%	97.47%
% Duplicate	0.05%	0.12%
% Mismatch	0.37%	34%
% soft clipped	2.07%	2.05%
% Q30 bases	90.59%	91.42%
Non-duplicate proper read pairs	769,150	803,473 (4.46% 34,323)

Figure 40

DENSE RUN 30pM Tile: 2103 (7 cycles)

	RTA	Within Boundary +/- 10 (Unet)
# clusters	1,157,285	1,351,588
% PF	78.58% (909,340)	71.62% (968,066)
% Aligned	97.69%	97.62%
% Duplicate	0.05%	0.07%
% Mismatch	0.65%	0.63%
% soft clipped	2.60%	2.57%
% Q30 bases	83.45%	83.88%
Non-duplicate proper read pairs	874,000	928,787 (6.27%)

Figure 41

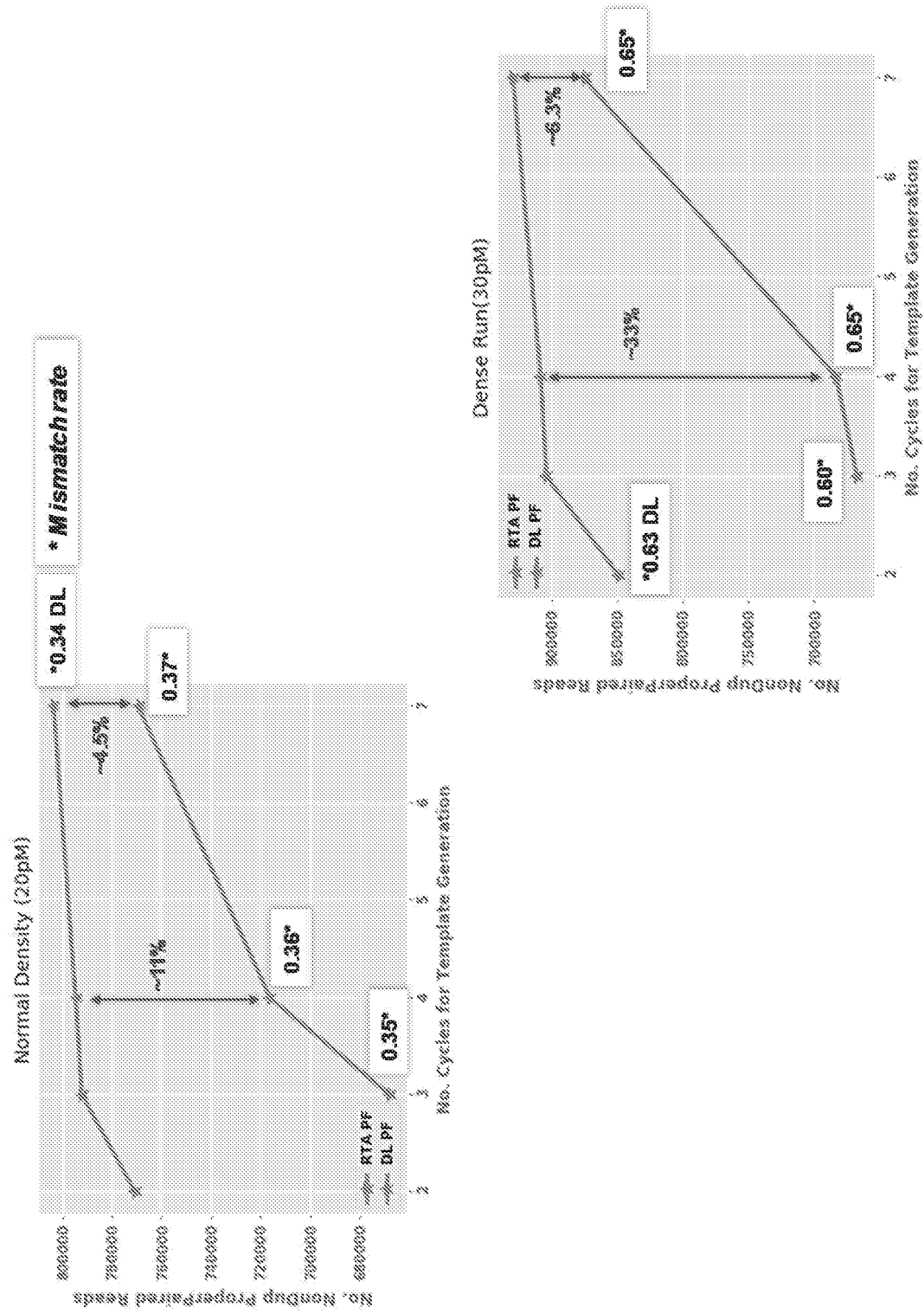
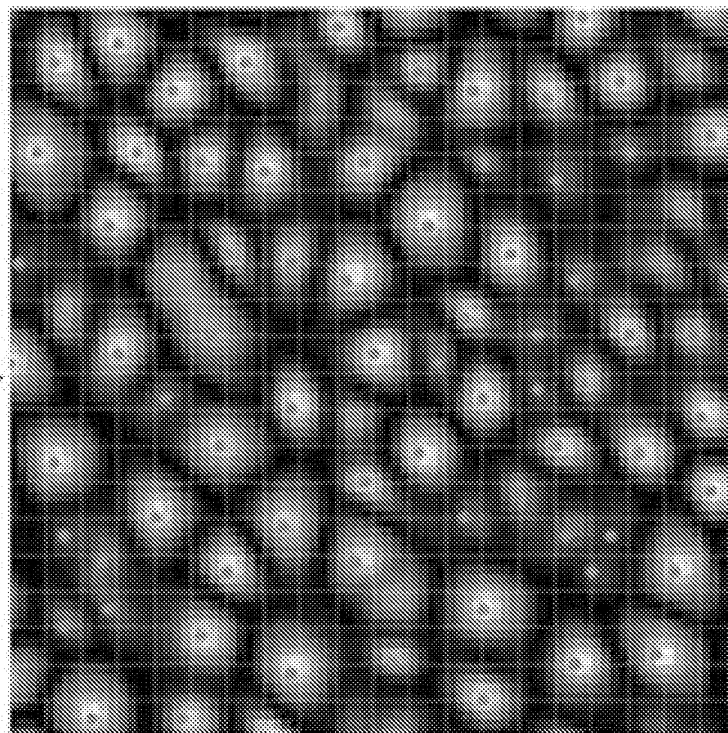


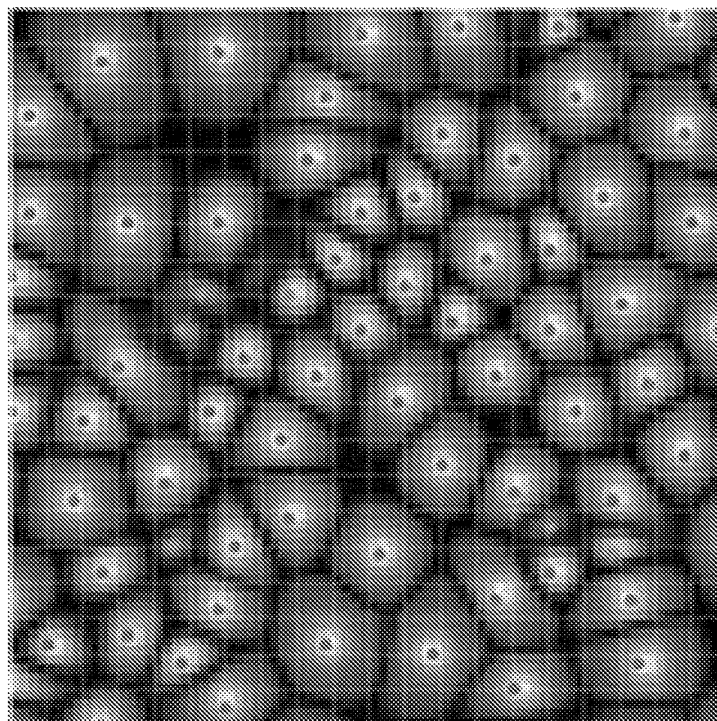
Figure 42

DENSE RUN 40pM Tile: 1101 (7 cycles)

40pM (Dense)



20pM (Normal)



□ Model identified smaller clusters

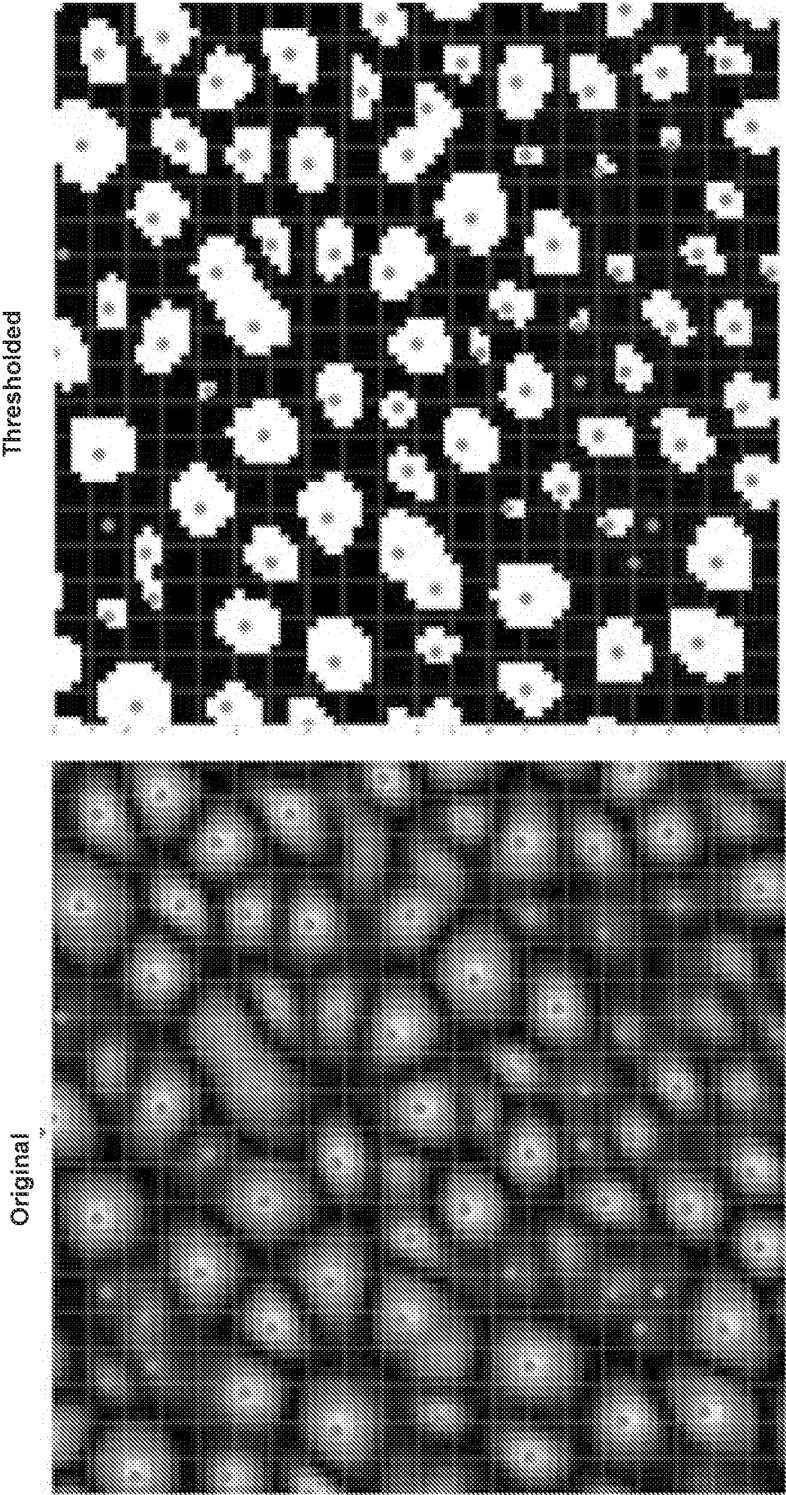
Figure 43

DENSE RUN 40pM Tile: 1101 (7 cycles)

	RTA	Within Boundary +/- 10 (Unet)
# clusters	1,089,460	1,509,395
% PF	42.60% (464,095)	54.00% (815,168)
% Aligned	93.00%	96.58%
% Duplicate	0.25%	0.04%
% Mismatch	2.93%	2.02%
% soft clipped	4.79%	11.83%
% Q30 bases	60.10%	65.33%
Non-duplicate proper read pairs	415,877	771,851 (85.6%?)
Total aligned bases	117,893,944	207,335,632

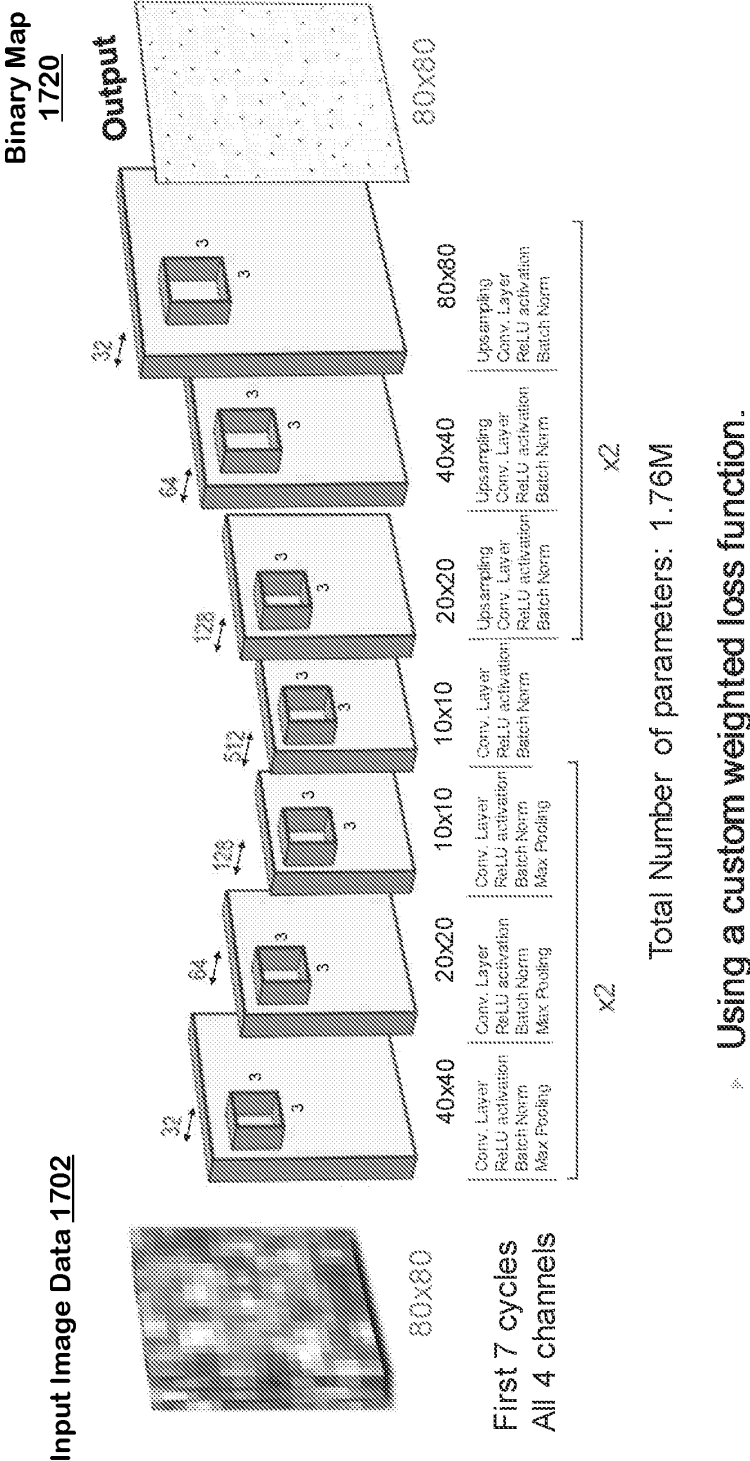
Figure 44

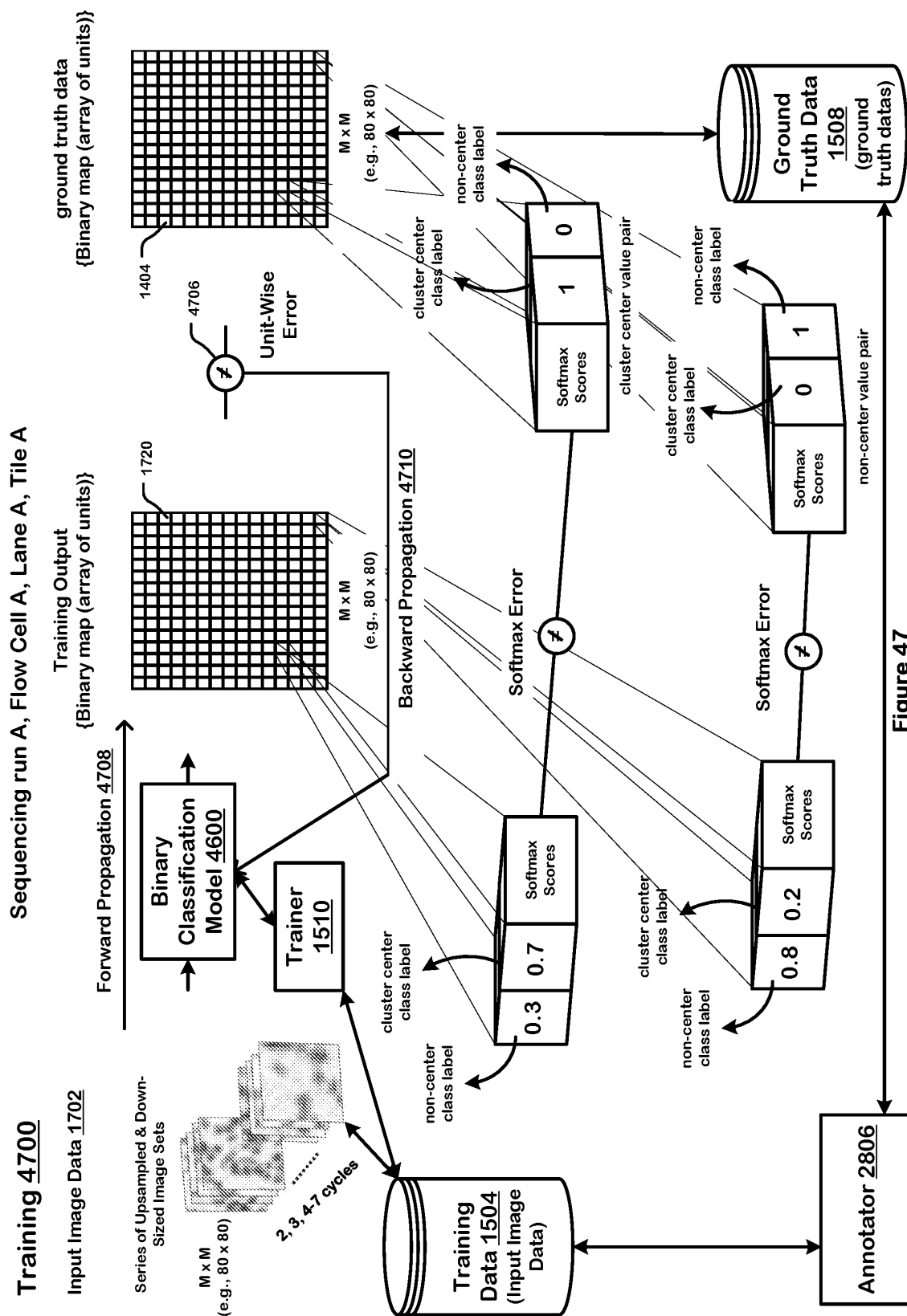
DENSE RUN 40pM Tile: 1101 (7 cycles)

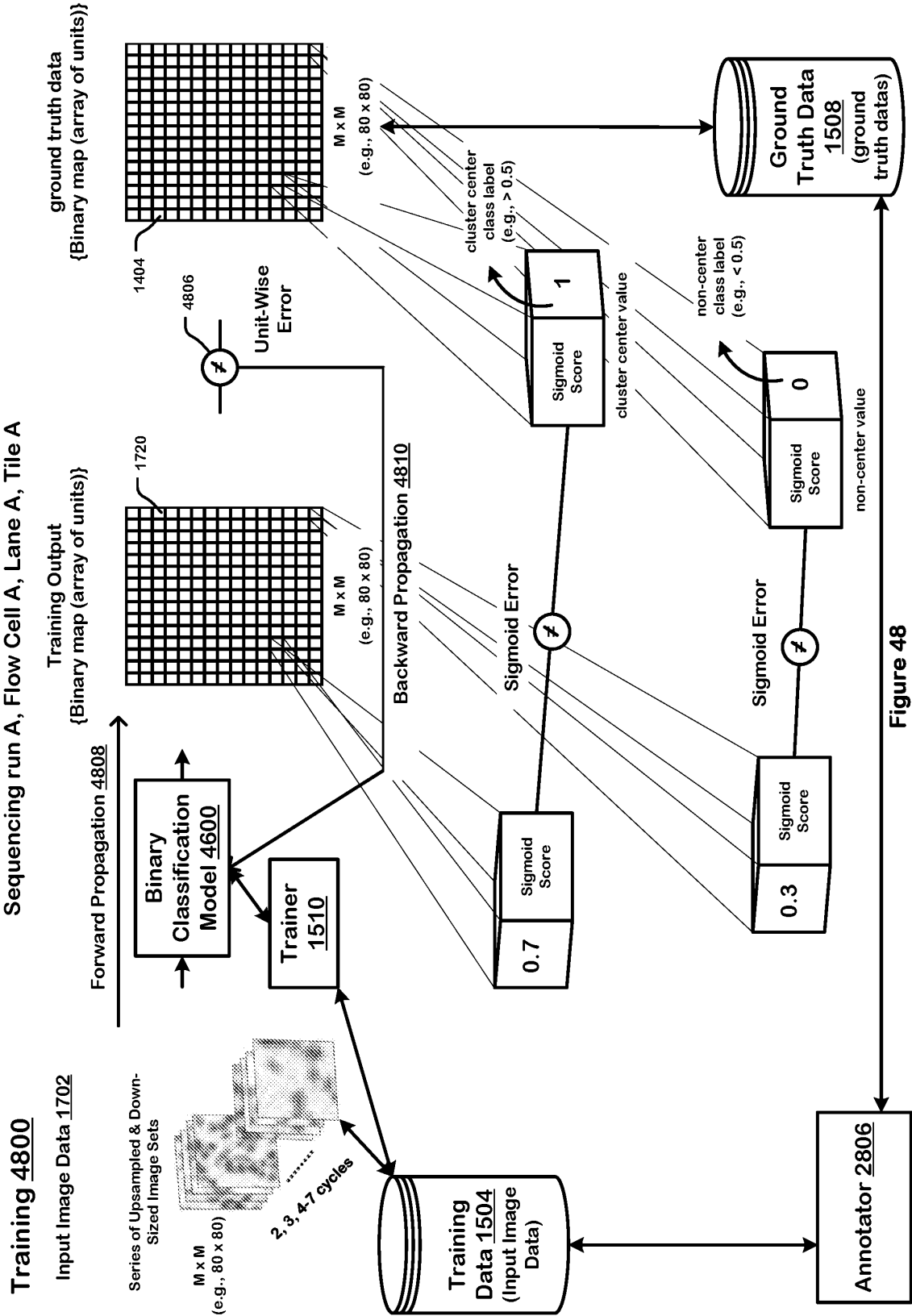


Identify cluster boundaries
Extract intensity
Improve mismatch rate using passing filter

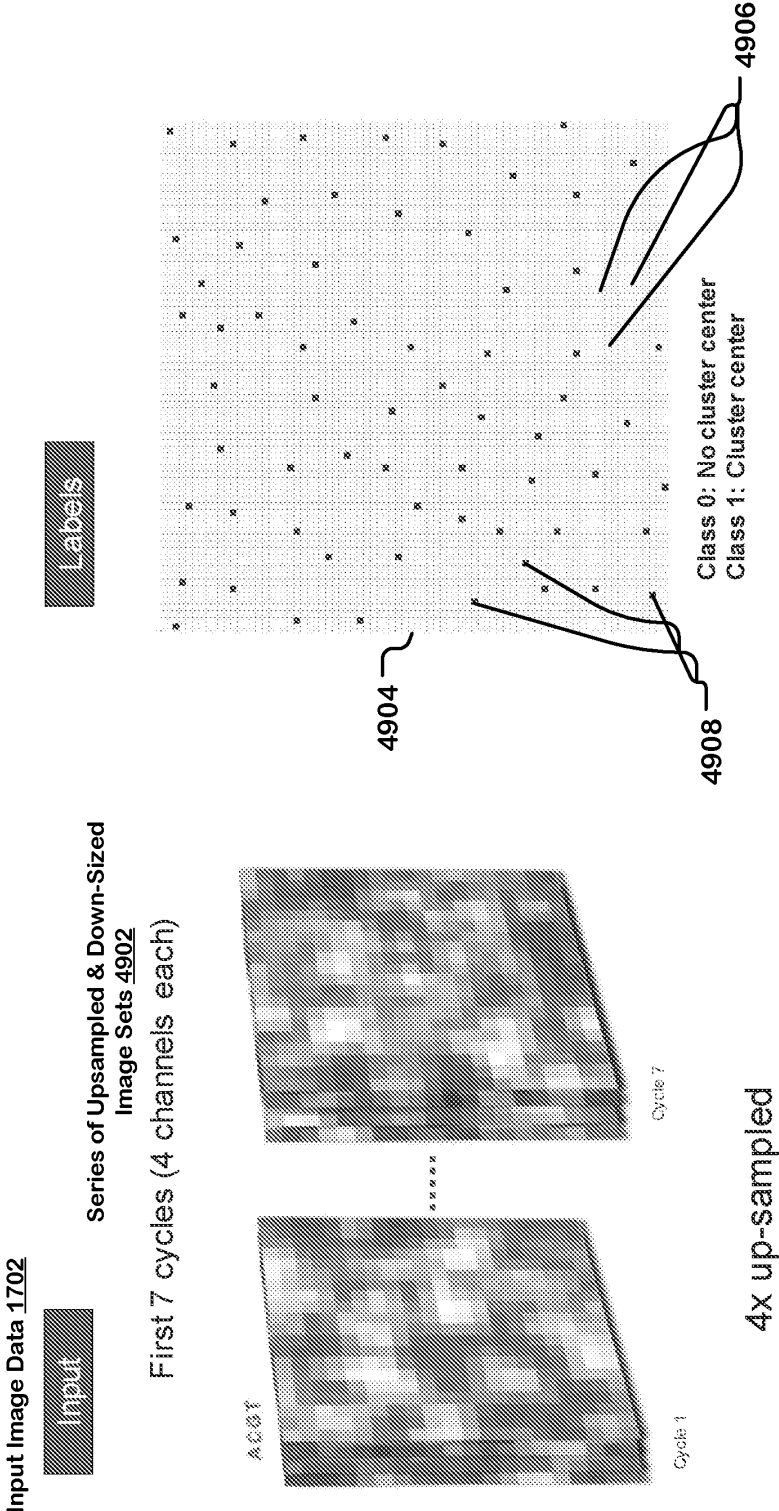
Figure 45







Binary Classification: Training Labels



Training labels: Cluster centers from ground truth data

Figure 49

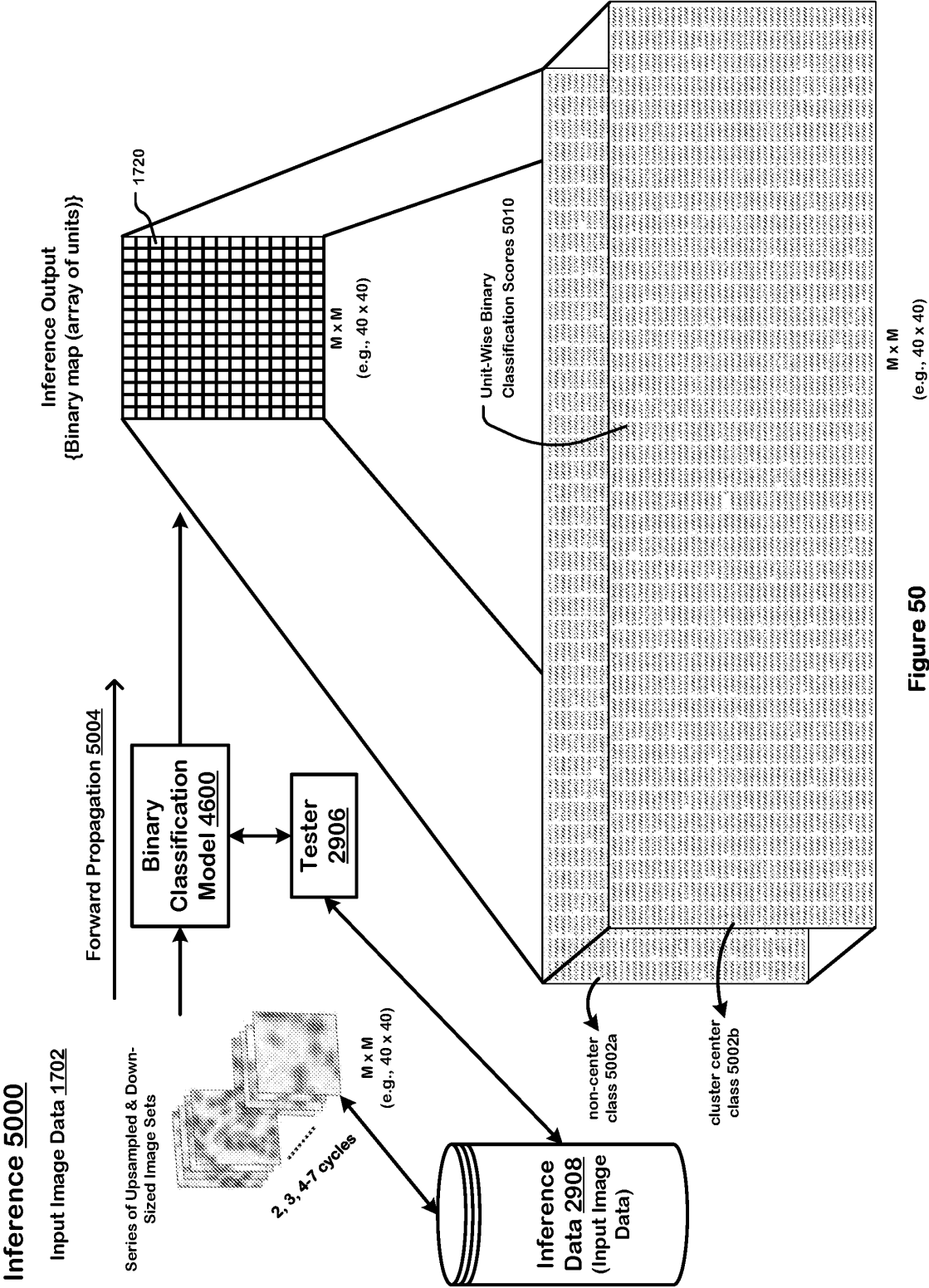


Figure 50

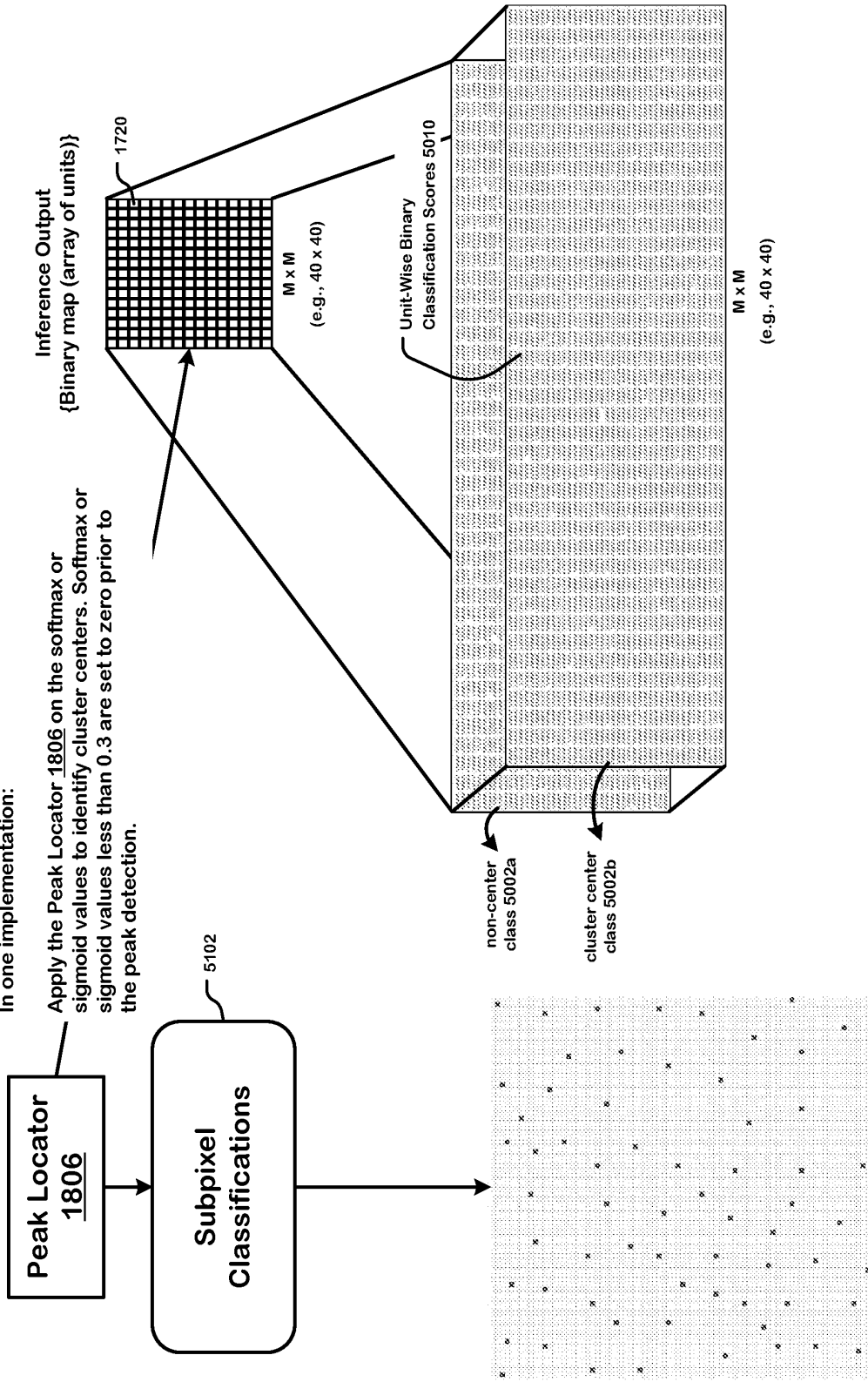


Figure 51

Binary Classification: Predictions

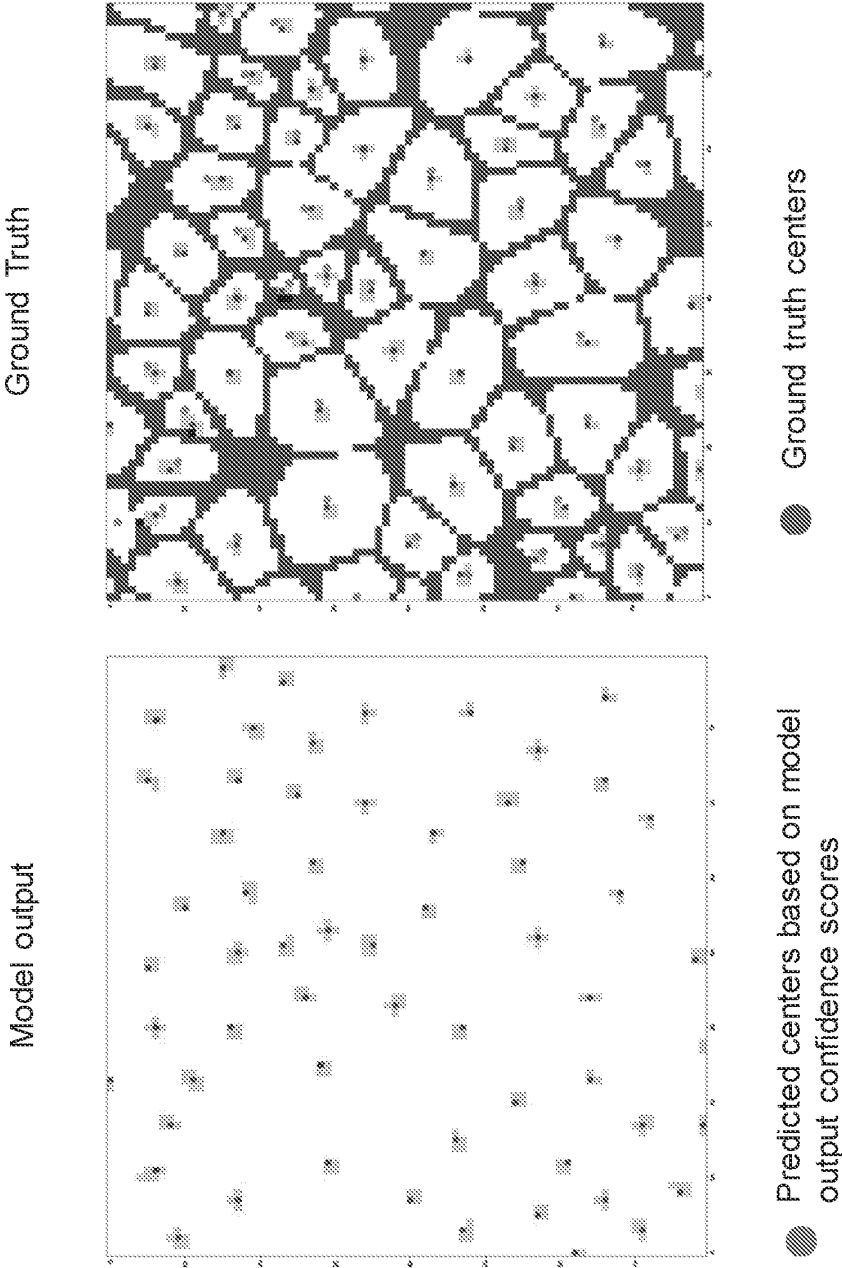


Figure 52a

Binary Classification: Predictions

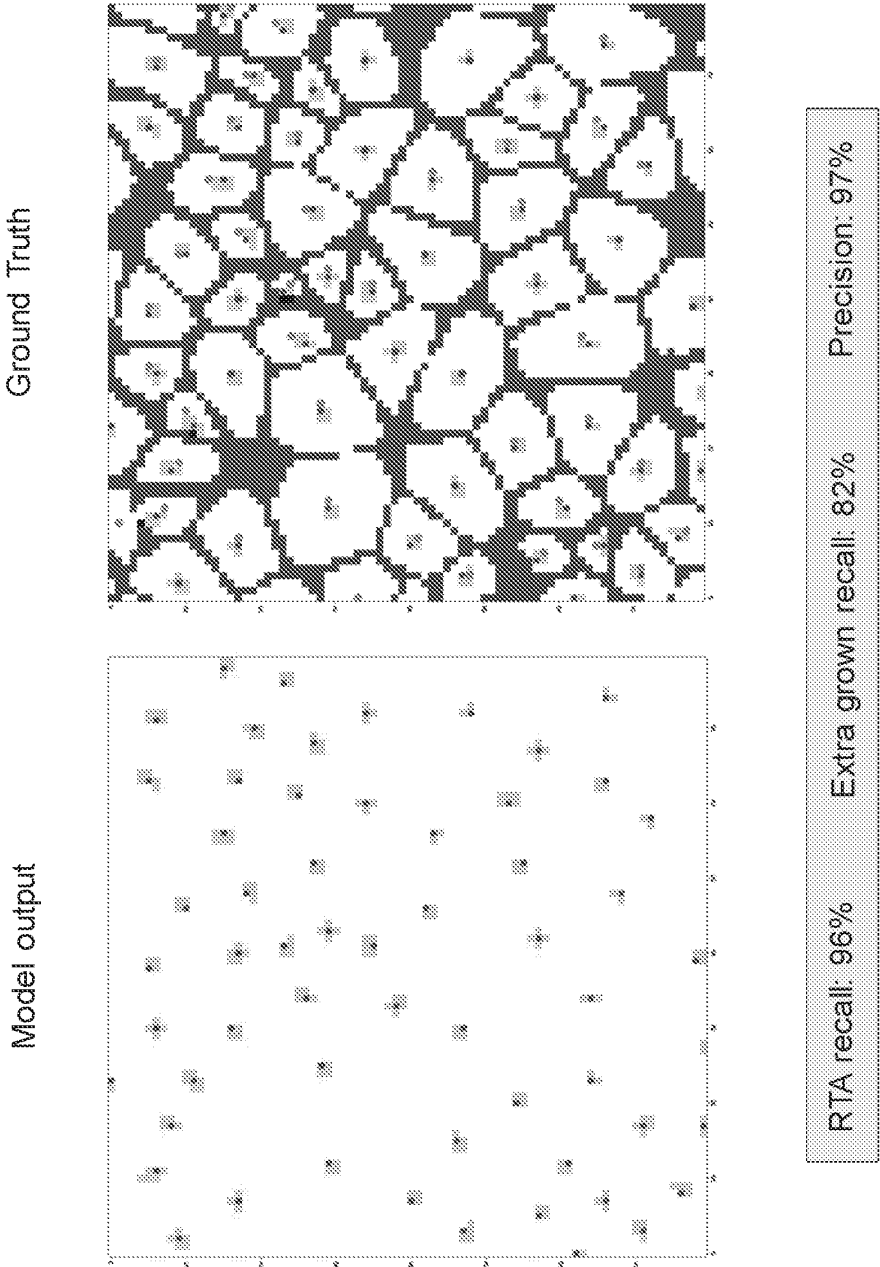
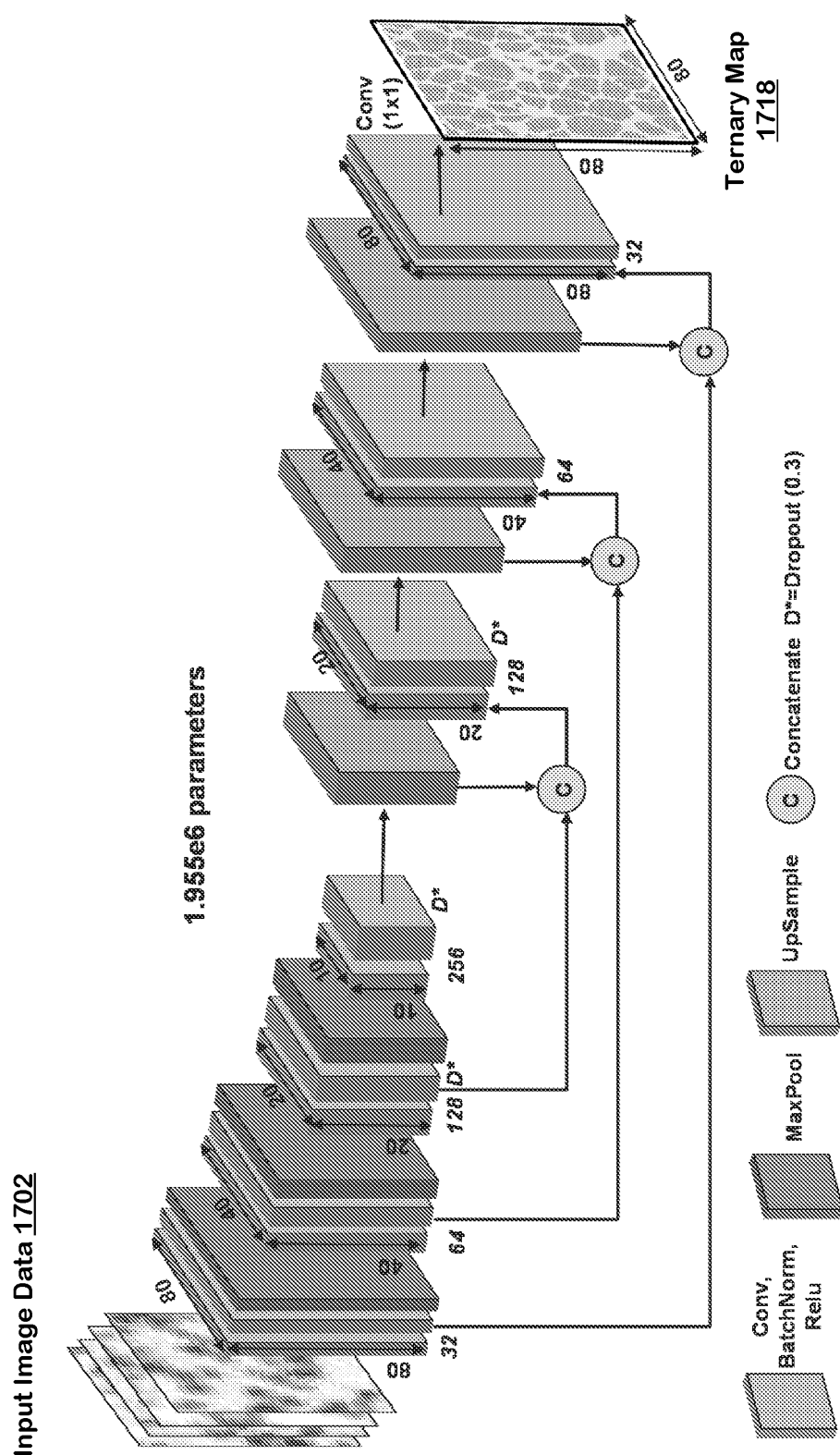


Figure 52b

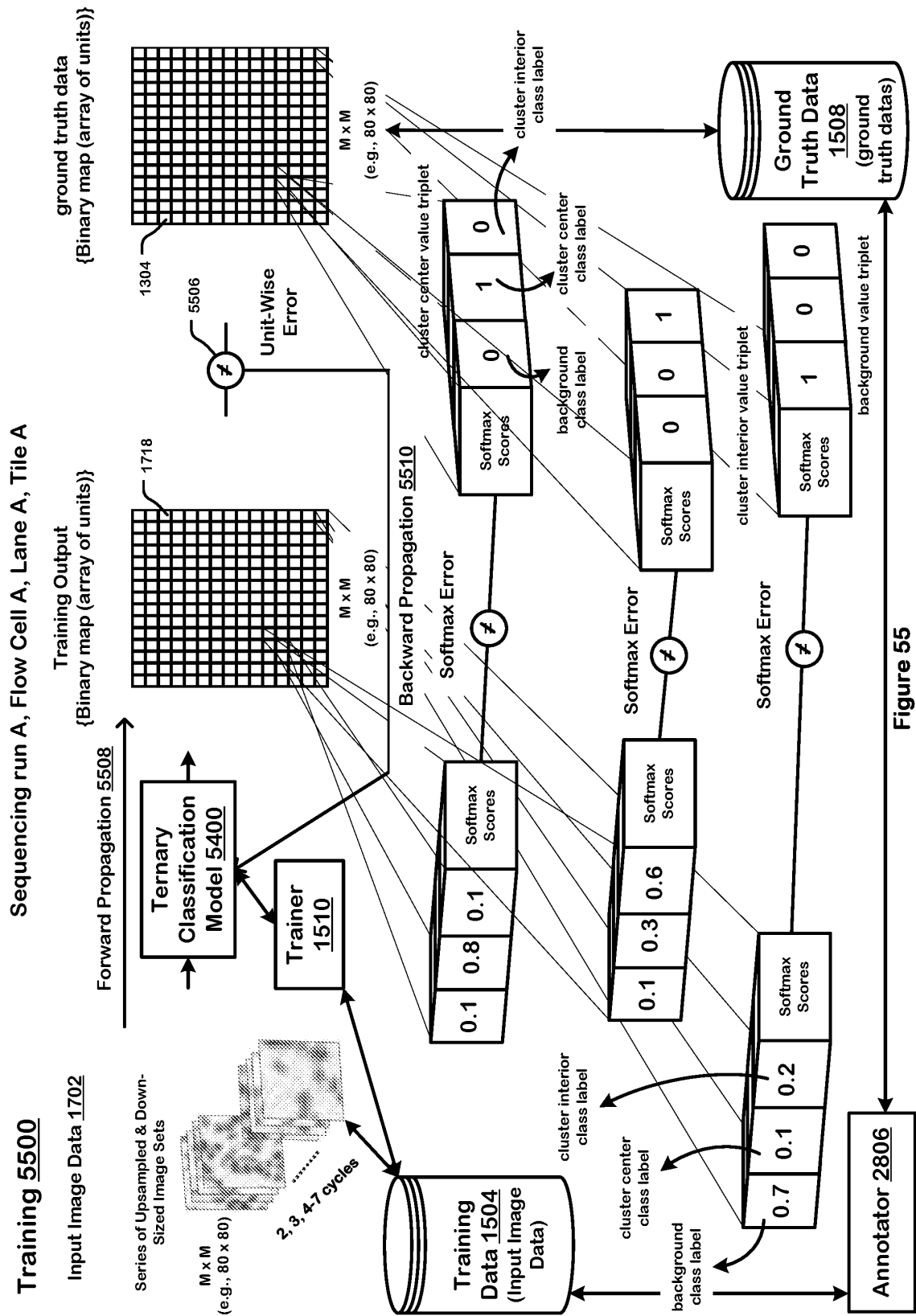
layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 80, 80, 28)	0
conv2d_1 (Conv2D)	(None, 80, 80, 32)	8064
batch_normalization_1 (Batch Normalization)	(None, 80, 80, 32)	128
conv2d_2 (Conv2D)	(None, 80, 80, 32)	9216
batch_normalization_2 (Batch Normalization)	(None, 80, 80, 32)	128
max_pooling2d_1 (MaxPooling2D)	(None, 40, 40, 32)	0
conv2d_3 (Conv2D)	(None, 40, 40, 64)	10432
batch_normalization_3 (Batch Normalization)	(None, 40, 40, 64)	256
conv2d_4 (Conv2D)	(None, 40, 40, 64)	36864
batch_normalization_4 (Batch Normalization)	(None, 40, 40, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None, 20, 20, 64)	0
conv2d_5 (Conv2D)	(None, 20, 20, 128)	73728
batch_normalization_5 (Batch Normalization)	(None, 20, 20, 128)	512
conv2d_6 (Conv2D)	(None, 20, 20, 128)	147456
batch_normalization_6 (Batch Normalization)	(None, 20, 20, 128)	512
dropout_1 (Dropout)	(None, 20, 20, 128)	0
max_pooling2d_3 (MaxPooling2D)	(None, 10, 10, 128)	0
conv2d_7 (Conv2D)	(None, 10, 10, 512)	588824
batch_normalization_7 (Batch Normalization)	(None, 10, 10, 512)	2048
dropout_2 (Dropout)	(None, 10, 10, 512)	0
up_sampling2d_1 (Upsampling2D)	(None, 20, 20, 512)	0
conv2d_8 (Conv2D)	(None, 40, 40, 128)	588824
batch_normalization_8 (Batch Normalization)	(None, 40, 40, 128)	512
conv2d_9 (Conv2D)	(None, 20, 20, 128)	147456
batch_normalization_9 (Batch Normalization)	(None, 20, 20, 128)	512
dropout_3 (Dropout)	(None, 20, 20, 128)	0
up_sampling2d_2 (Upsampling2D)	(None, 40, 40, 128)	0
conv2d_10 (Conv2D)	(None, 40, 40, 64)	73728
batch_normalization_10 (Batch Normalization)	(None, 40, 40, 64)	256
conv2d_11 (Conv2D)	(None, 40, 40, 64)	36864
batch_normalization_11 (Batch Normalization)	(None, 40, 40, 64)	256
up_sampling2d_3 (Upsampling2D)	(None, 80, 80, 64)	0
conv2d_12 (Conv2D)	(None, 80, 80, 32)	18432
batch_normalization_12 (Batch Normalization)	(None, 80, 80, 32)	128
conv2d_13 (Conv2D)	(None, 80, 80, 32)	9216
batch_normalization_13 (Batch Normalization)	(None, 80, 80, 32)	128
pred (Conv2D)	(None, 80, 80, 2)	64
Total params: 1,764,800		
Trainable params: 1,751,984		
Non-trainable params: 2,816		

Figure 53



Ternary Classification Model 5400

Figure 54



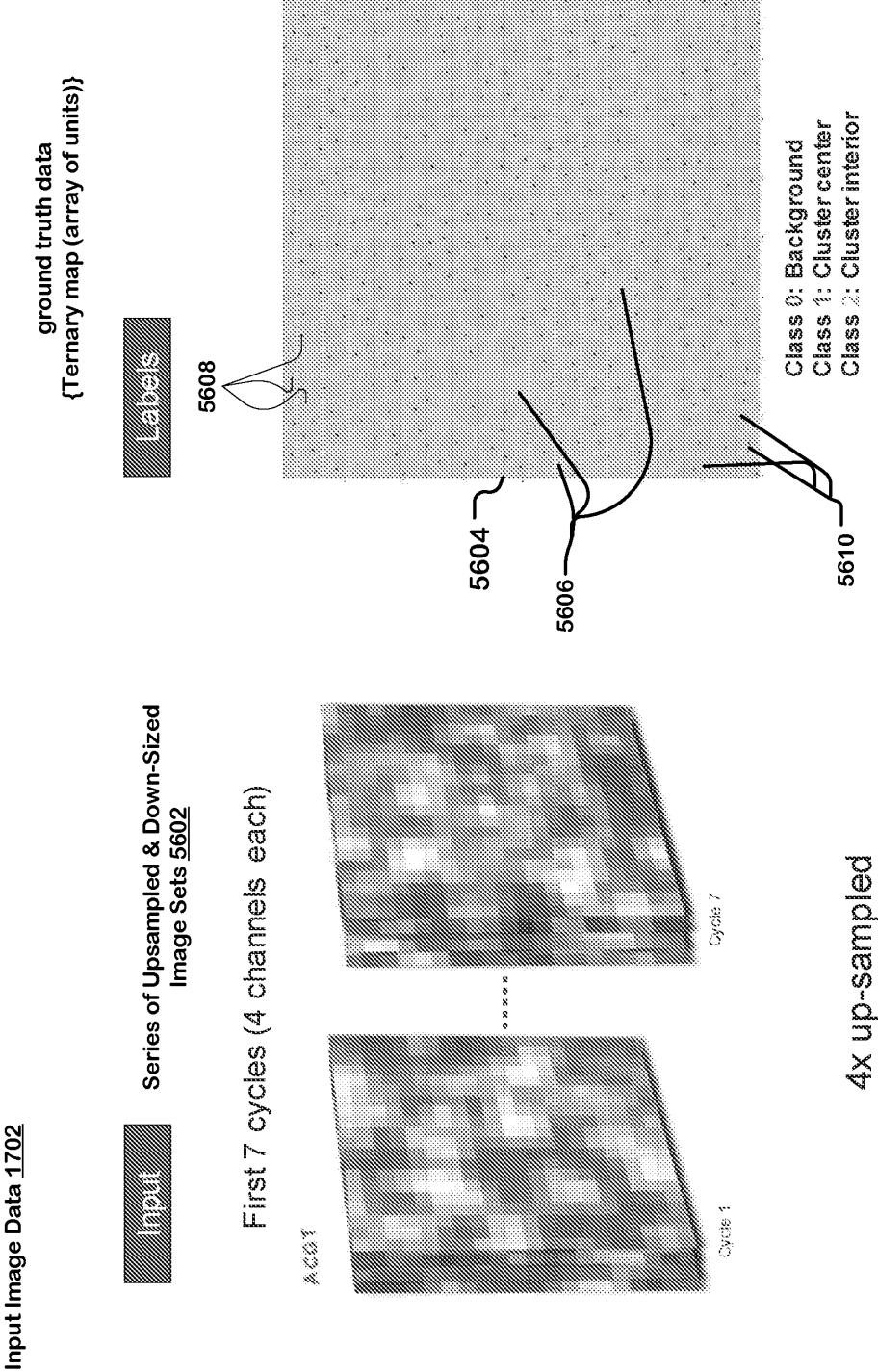


Figure 56

layer (type)	Output Shape	Params	Connected to
input_1 (InputLayer)	(None, None, None, 2, 0)		
batch_normalization_1 (BatchNormalizer)	(None, None, None, 2, 112)		input_1[0][0]
conv2d_1 (Conv2D)	(None, None, None, 3, 8064)		batch_normalization_1[0][0]
batch_normalization_2 (BatchNormalizer)	(None, None, None, 3, 128)		conv2d_1[0][0]
conv2d_2 (Conv2D)	(None, None, None, 3, 8216)		batch_normalization_2[0][0]
batch_normalization_3 (BatchNormalizer)	(None, None, None, 3, 128)		conv2d_2[0][0]
max_pooling2d_1 (MaxPooling2D)	(None, None, None, 1, 0)		batch_normalization_3[0][0]
conv2d_3 (Conv2D)	(None, None, None, 6, 16432)		max_pooling2d_1[0][0]
batch_normalization_4 (BatchNormalizer)	(None, None, None, 6, 256)		conv2d_3[0][0]
conv2d_4 (Conv2D)	(None, None, None, 6, 16064)		batch_normalization_4[0][0]
batch_normalization_5 (BatchNormalizer)	(None, None, None, 6, 256)		conv2d_4[0][0]
max_pooling2d_2 (MaxPooling2D)	(None, None, None, 8, 0)		batch_normalization_5[0][0]
conv2d_5 (Conv2D)	(None, None, None, 1, 73736)		max_pooling2d_2[0][0]
batch_normalization_6 (BatchNormalizer)	(None, None, None, 1, 512)		conv2d_5[0][0]
conv2d_6 (Conv2D)	(None, None, None, 1, 147456)		batch_normalization_6[0][0]
batch_normalization_7 (BatchNormalizer)	(None, None, None, 1, 512)		conv2d_6[0][0]
dropout_1 (Dropout)	(None, None, None, 1, 0)		batch_normalization_7[0][0]
max_pooling2d_3 (MaxPooling2D)	(None, None, None, 1, 0)		dropout_1[0][0]
conv2d_7 (Conv2D)	(None, None, None, 2, 254512)		max_pooling2d_3[0][0]
batch_normalization_8 (BatchNormalizer)	(None, None, None, 2, 1924)		conv2d_7[0][0]
conv2d_8 (Conv2D)	(None, None, None, 2, 743424)		batch_normalization_8[0][0]
batch_normalization_9 (BatchNormalizer)	(None, None, None, 2, 1924)		conv2d_8[0][0]
dropout_2 (Dropout)	(None, None, None, 2, 0)		batch_normalization_9[0][0]
up_sampling2d_1 (UpSampling2D)	(None, None, None, 2, 0)		dropout_2[0][0]
concatenate_1 (Concatenate)	(None, None, None, 3, 0)		up_sampling2d_1[0][0] batch_normalization_9[0][0]
conv2d_9 (Conv2D)	(None, None, None, 1, 442368)		concatenate_1[0][0]
batch_normalization_10 (BatchNormalizer)	(None, None, None, 1, 512)		conv2d_9[0][0]
conv2d_10 (Conv2D)	(None, None, None, 1, 147456)		batch_normalization_10[0][0]
batch_normalization_11 (BatchNormalizer)	(None, None, None, 1, 512)		conv2d_10[0][0]
dropout_3 (Dropout)	(None, None, None, 1, 0)		batch_normalization_11[0][0]
up_sampling2d_2 (UpSampling2D)	(None, None, None, 1, 0)		dropout_3[0][0]
concatenate_2 (Concatenate)	(None, None, None, 1, 0)		up_sampling2d_2[0][0] batch_normalization_11[0][0]
conv2d_11 (Conv2D)	(None, None, None, 6, 110092)		concatenate_2[0][0]
batch_normalization_12 (BatchNormalizer)	(None, None, None, 6, 256)		conv2d_11[0][0]
conv2d_12 (Conv2D)	(None, None, None, 6, 16064)		batch_normalization_12[0][0]
batch_normalization_13 (BatchNormalizer)	(None, None, None, 6, 256)		conv2d_12[0][0]
up_sampling2d_3 (UpSampling2D)	(None, None, None, 8, 0)		batch_normalization_13[0][0]
concatenate_3 (Concatenate)	(None, None, None, 9, 0)		up_sampling2d_3[0][0] batch_normalization_13[0][0]
conv2d_13 (Conv2D)	(None, None, None, 3, 27128)		concatenate_3[0][0]
batch_normalization_14 (BatchNormalizer)	(None, None, None, 3, 128)		conv2d_13[0][0]
conv2d_14 (Conv2D)	(None, None, None, 3, 8216)		batch_normalization_14[0][0]
batch_normalization_15 (BatchNormalizer)	(None, None, None, 3, 128)		conv2d_14[0][0]
global_avg_pooling2d (GlobalAveragePooling2D)	(None, None, None, 3, 36)		batch_normalization_15[0][0]
Total params: 1,556,480			
Trainable params: 1,556,480			
Non-trainable params: 0,000			

Figure 57

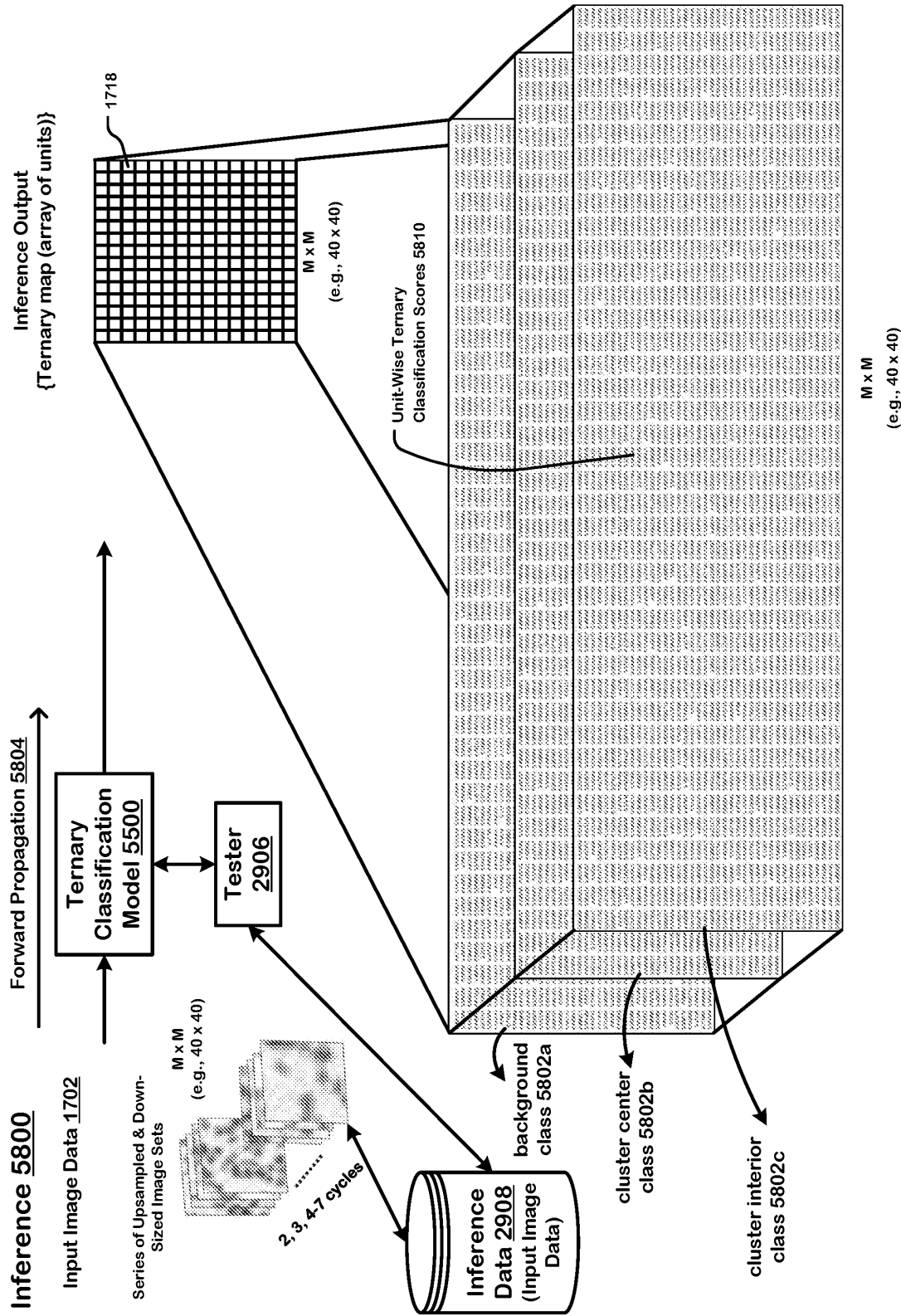
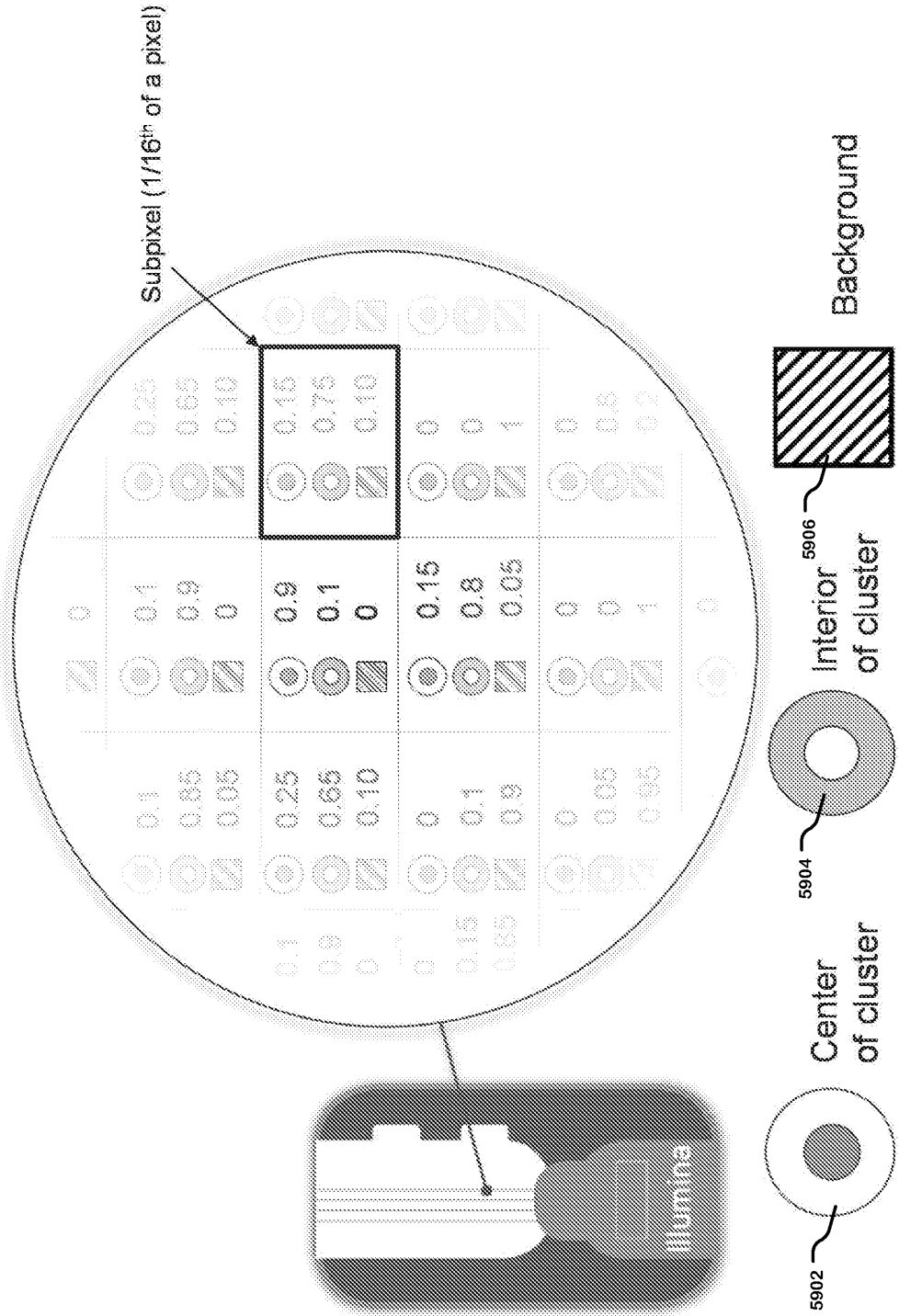
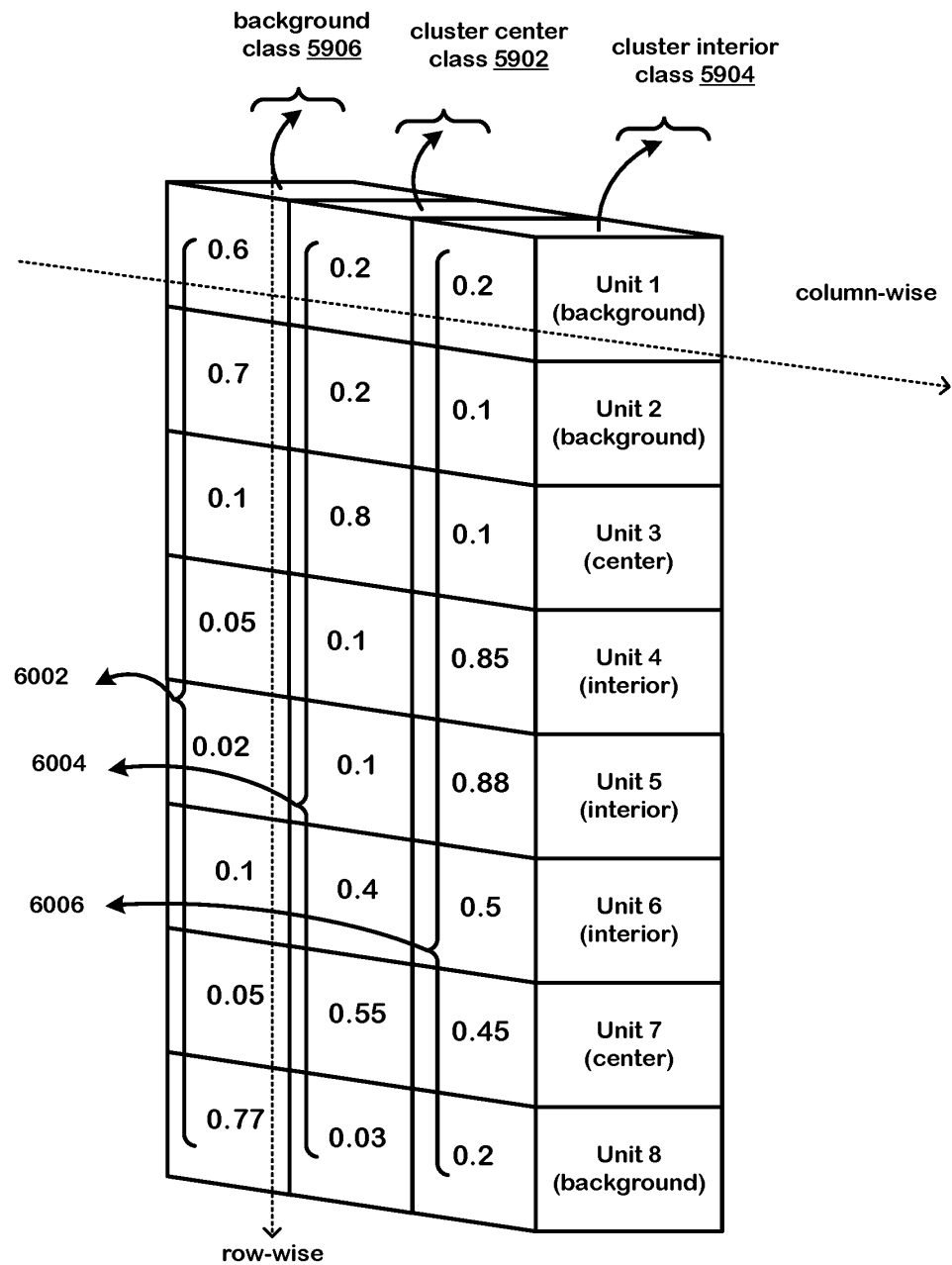


Figure 58

Prediction scores for every subpixel





Array of Units

Figure 60

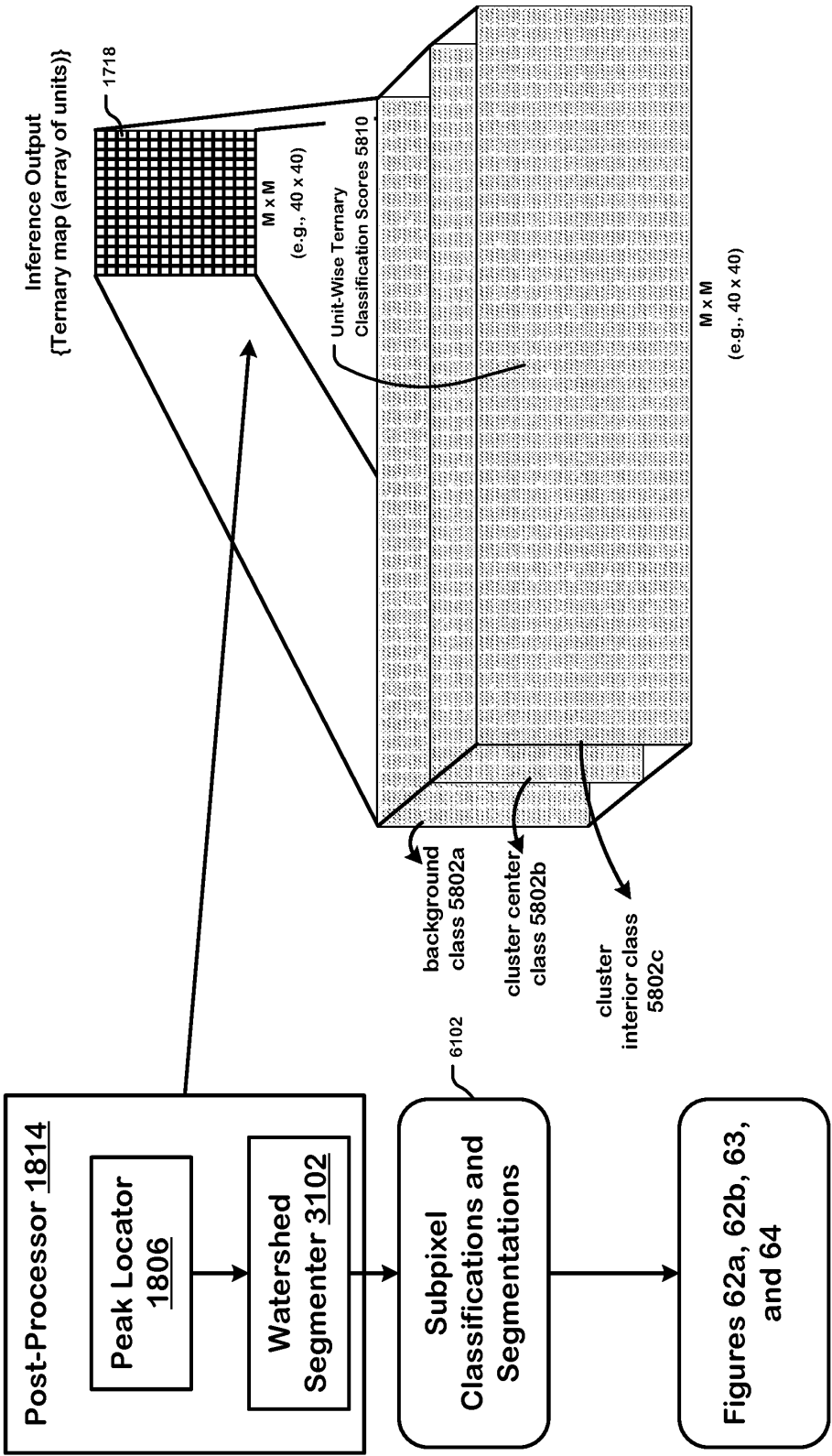
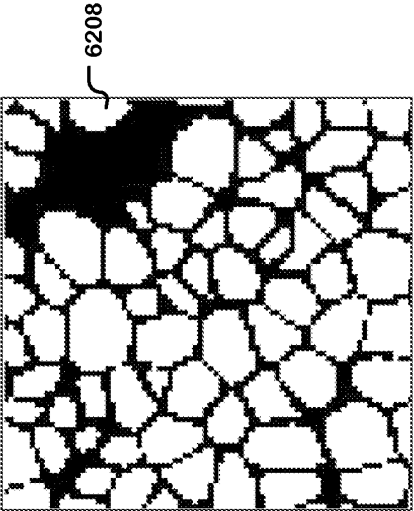
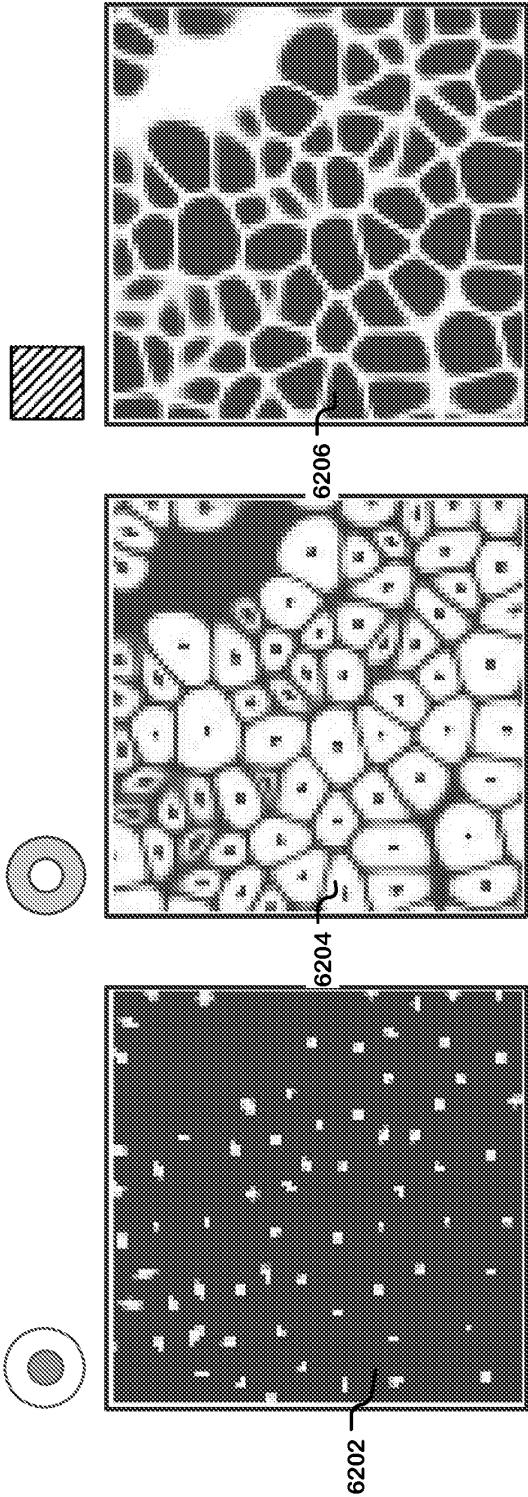


Figure 61

Model Predictions



Ground truth:

Figure 62a

Three-Category Classification: Predictions

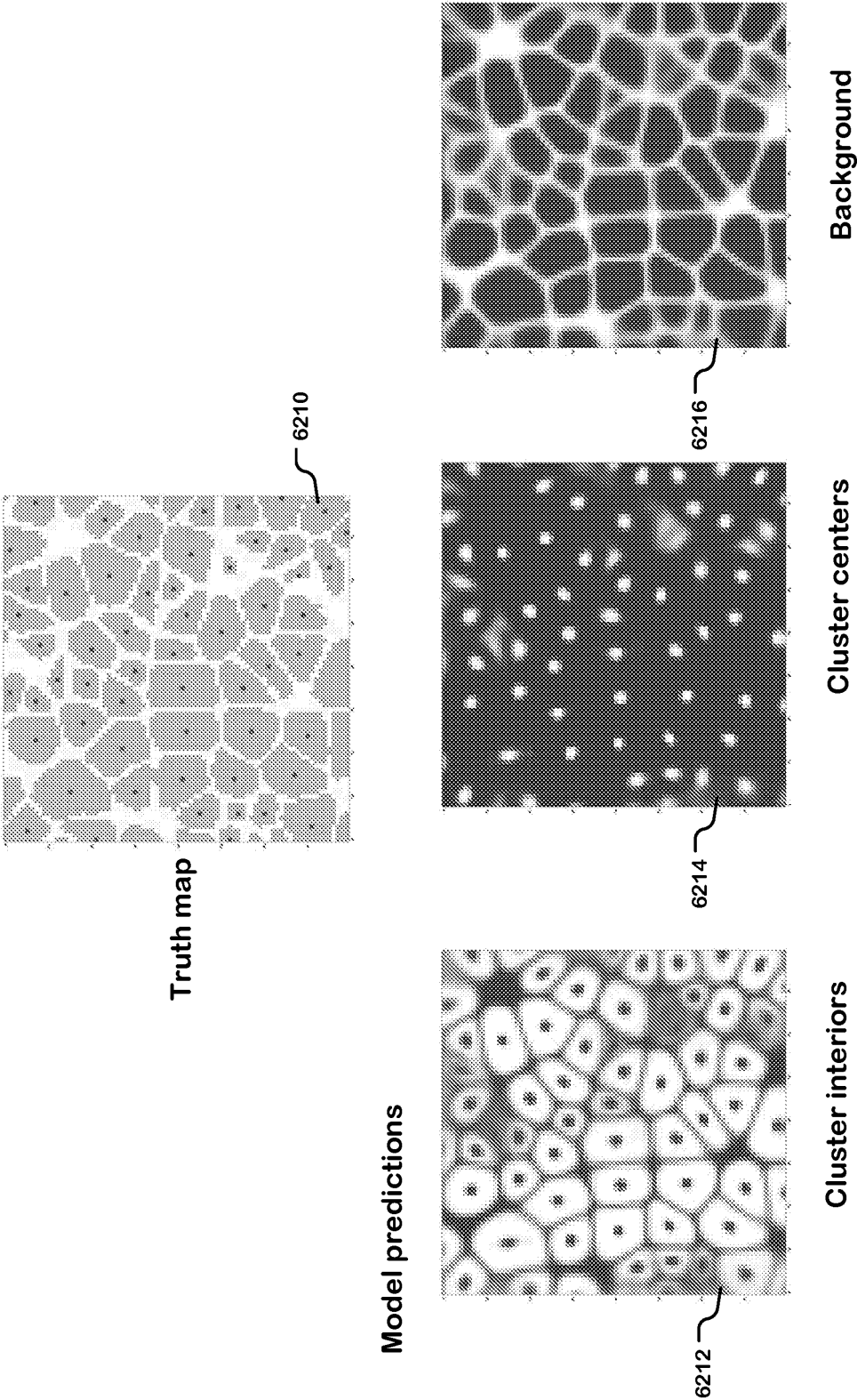


Figure 62b

Three-Category Classification: Predictions

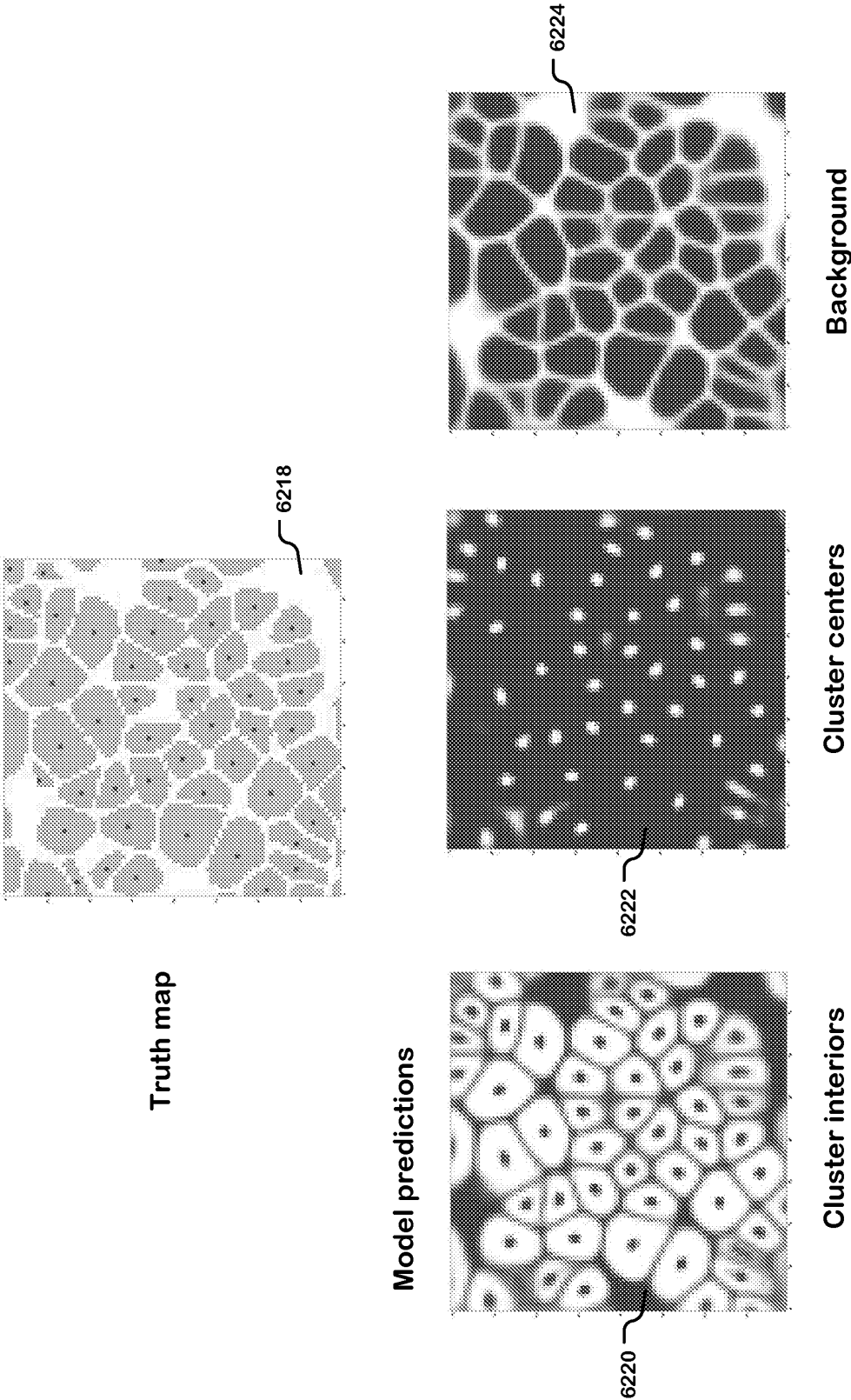


Figure 62c

Identifying Cluster Shape Data

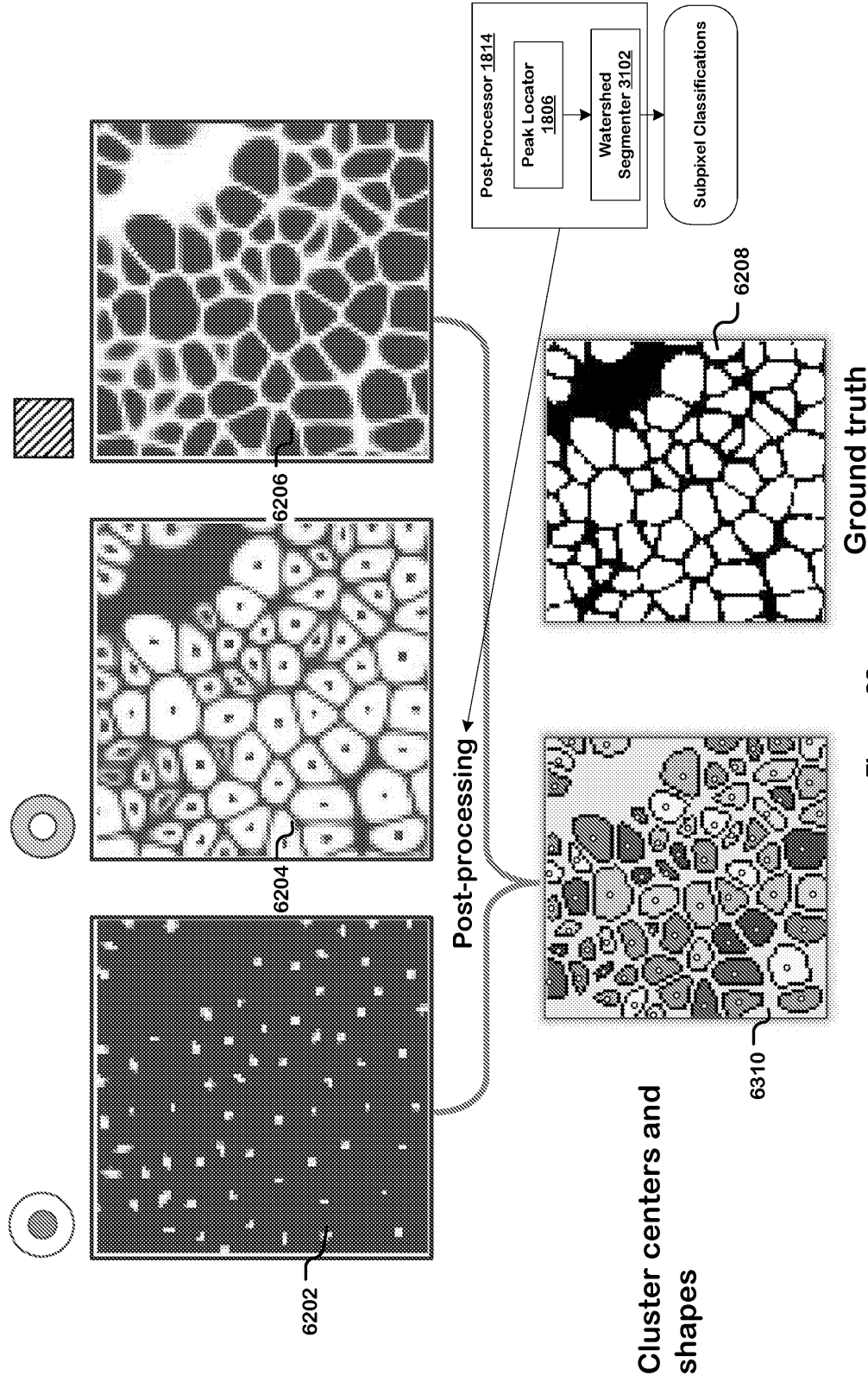


Figure 63

Basecall comparisons (using RTA 1.18)

	RTA	DL binary classification	DL regression
# clusters	493,944	519,100	543,348
% PF	88.8%	86.99%	85.86%
% Aligned	97.63%	97.68%	97.67%
% Duplicate	0.02%	0.2%	0.14%
% Mismatch	0.34%	0.29%	0.29%
% soft clipped	2.02%	1.97%	1.97%
% Q30 bases	91.94%	92.76%	92.81%
Total proper read pairs	422,907	435,737	449,993
Non-duplicate proper read pairs	424,665	434,869	449,344

Figure 64

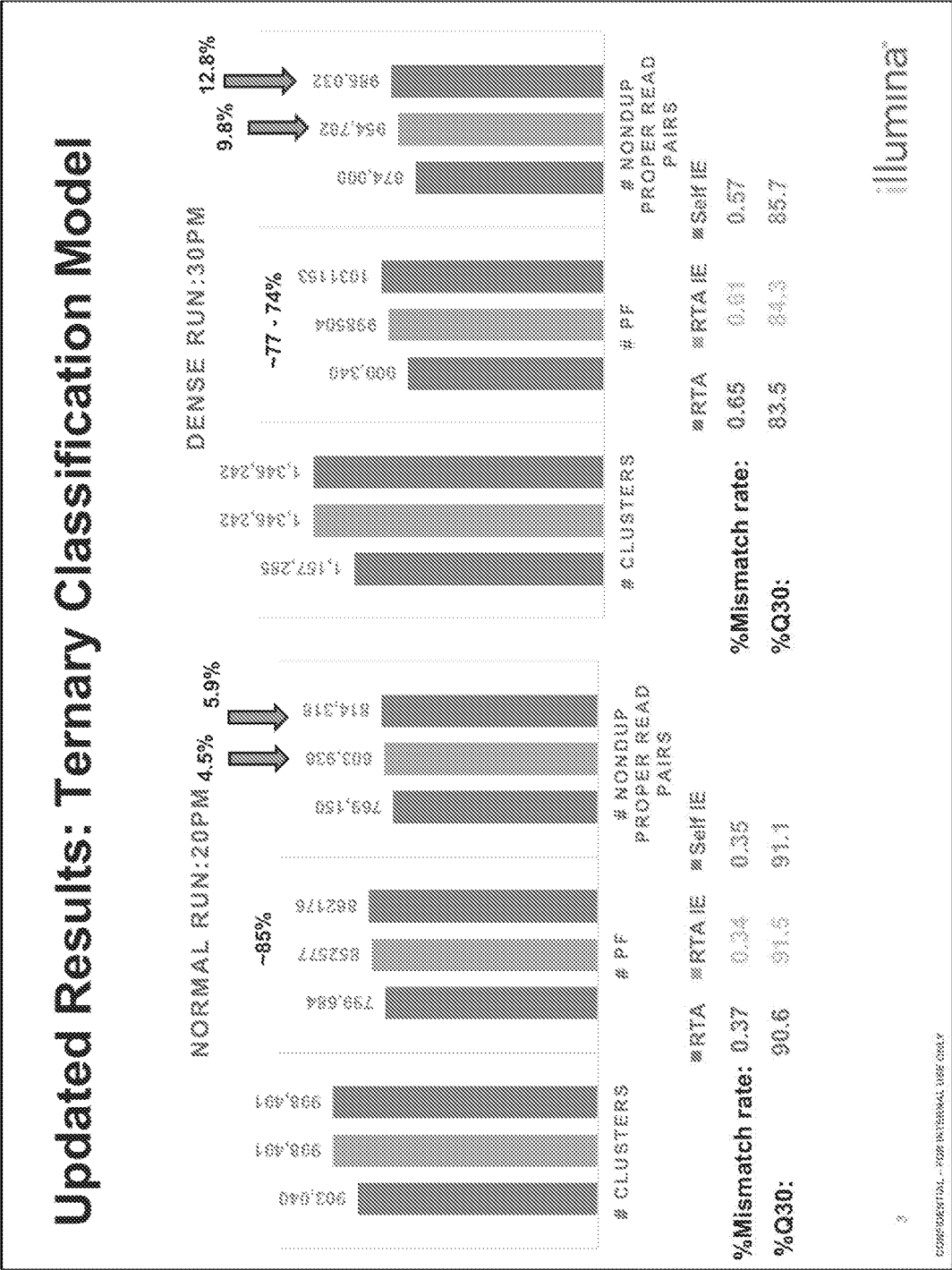


Figure 65

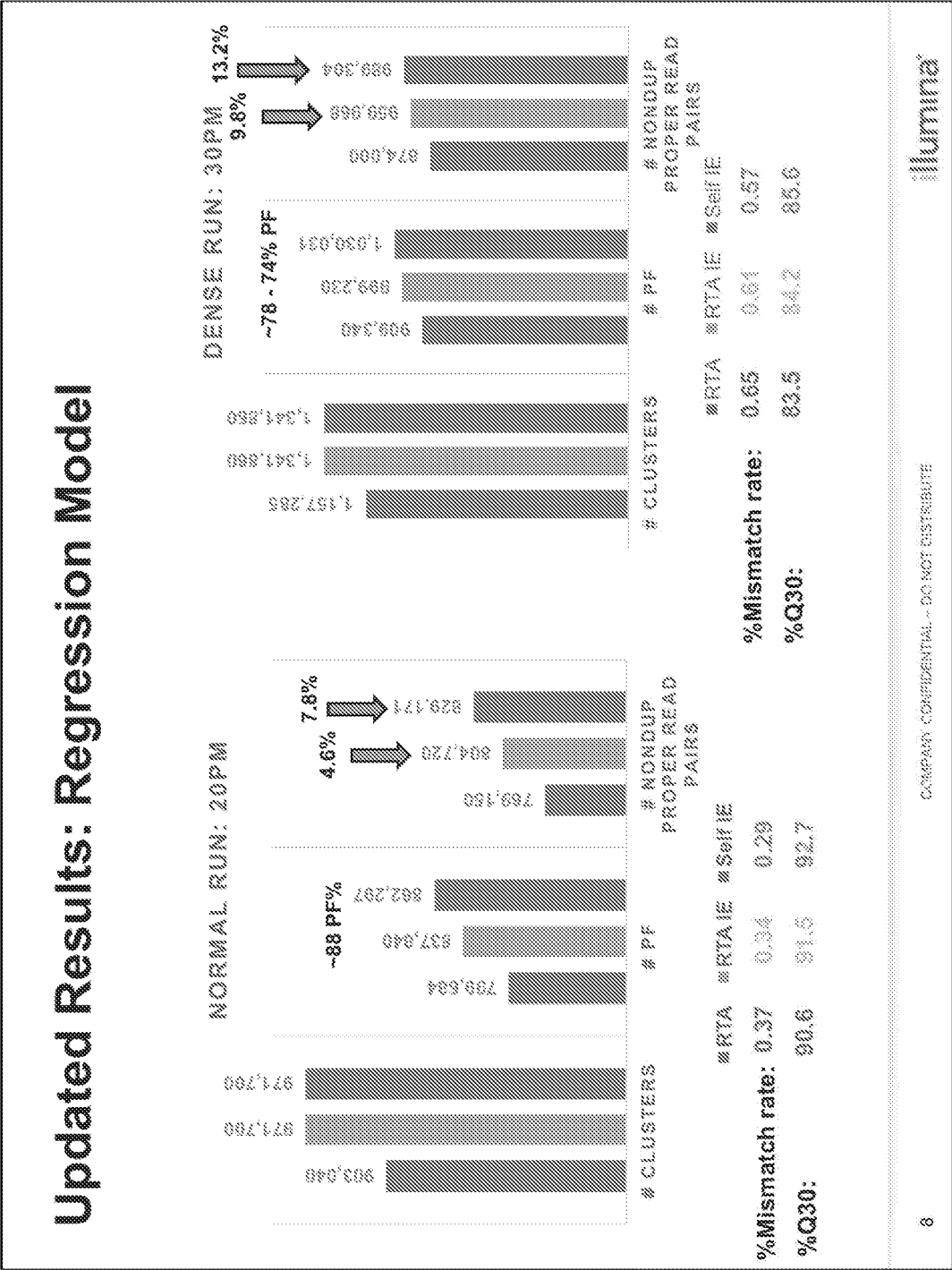


Figure 66

What has the model learned?

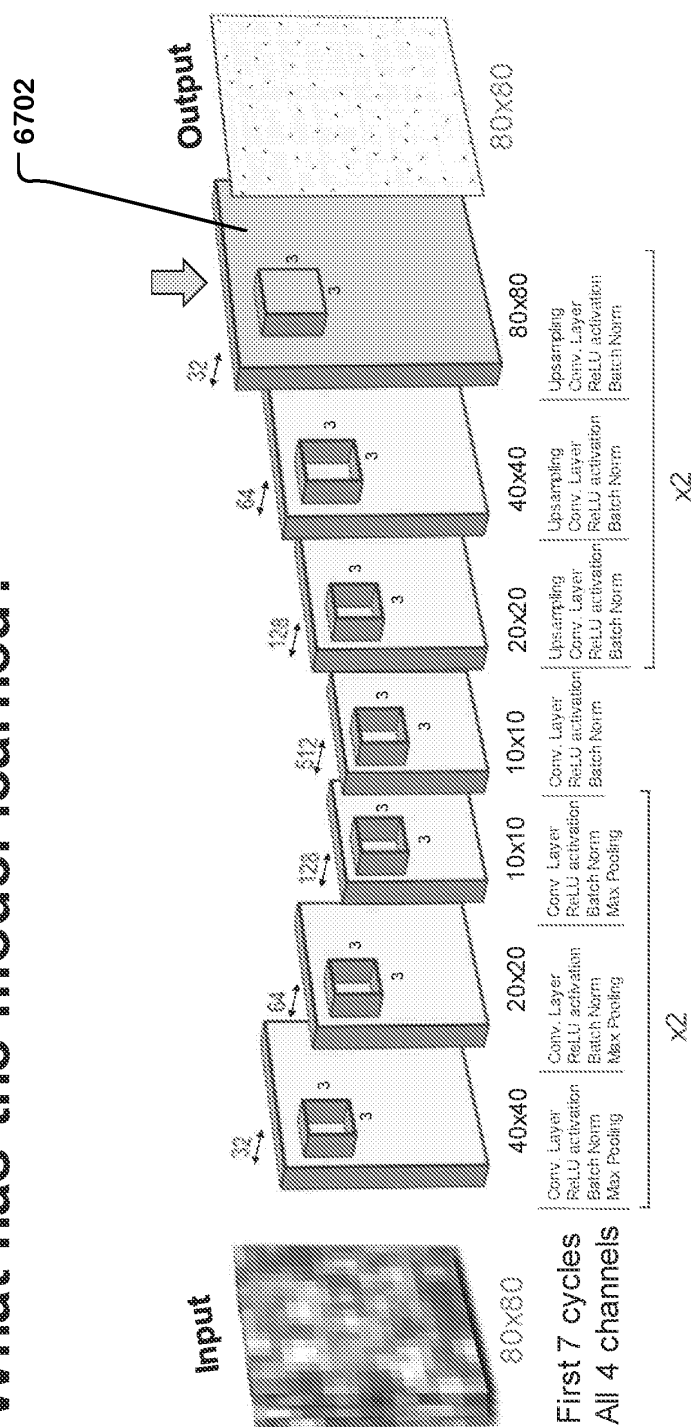


Figure 67

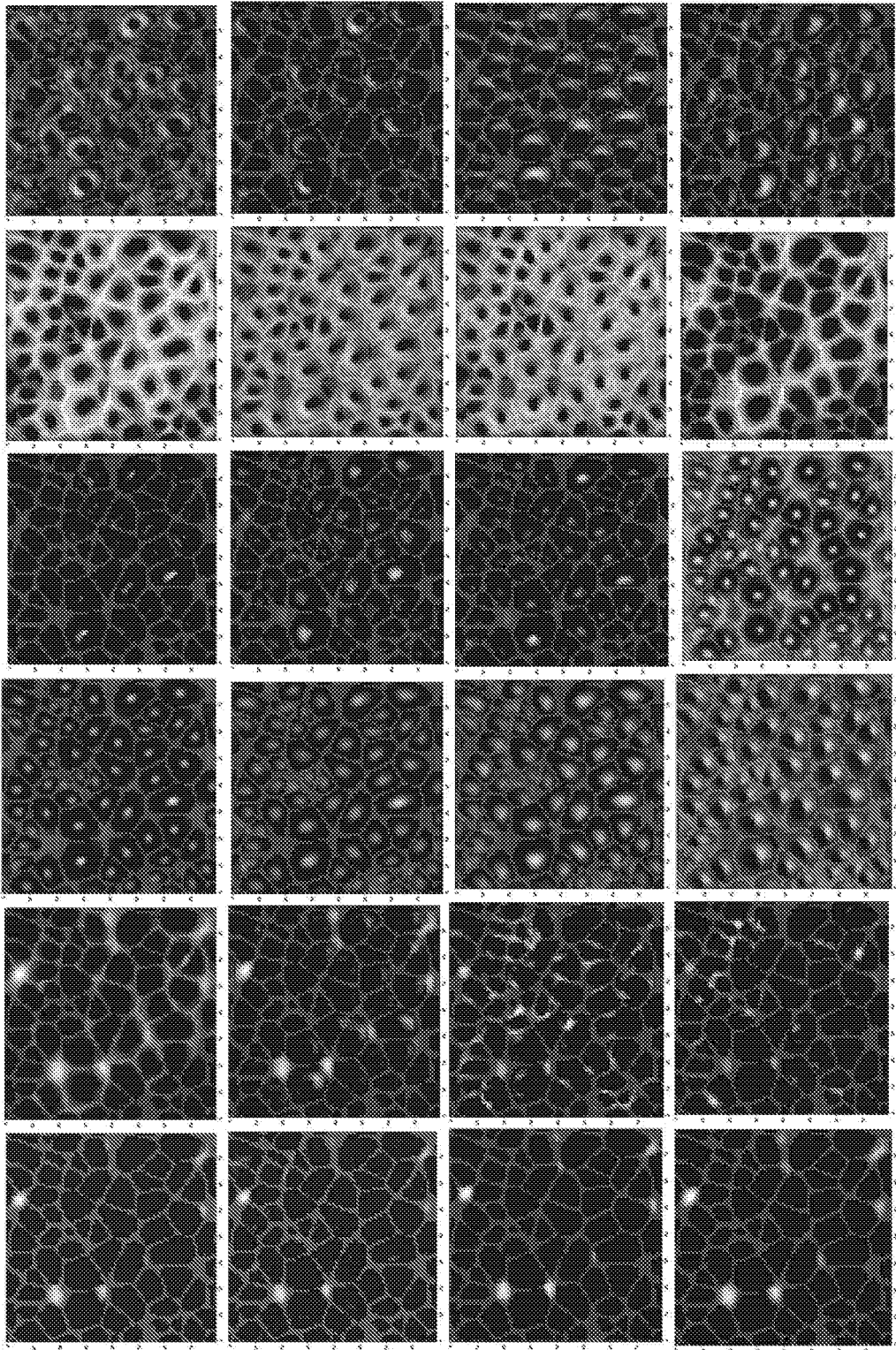


Figure 68

NextSeq

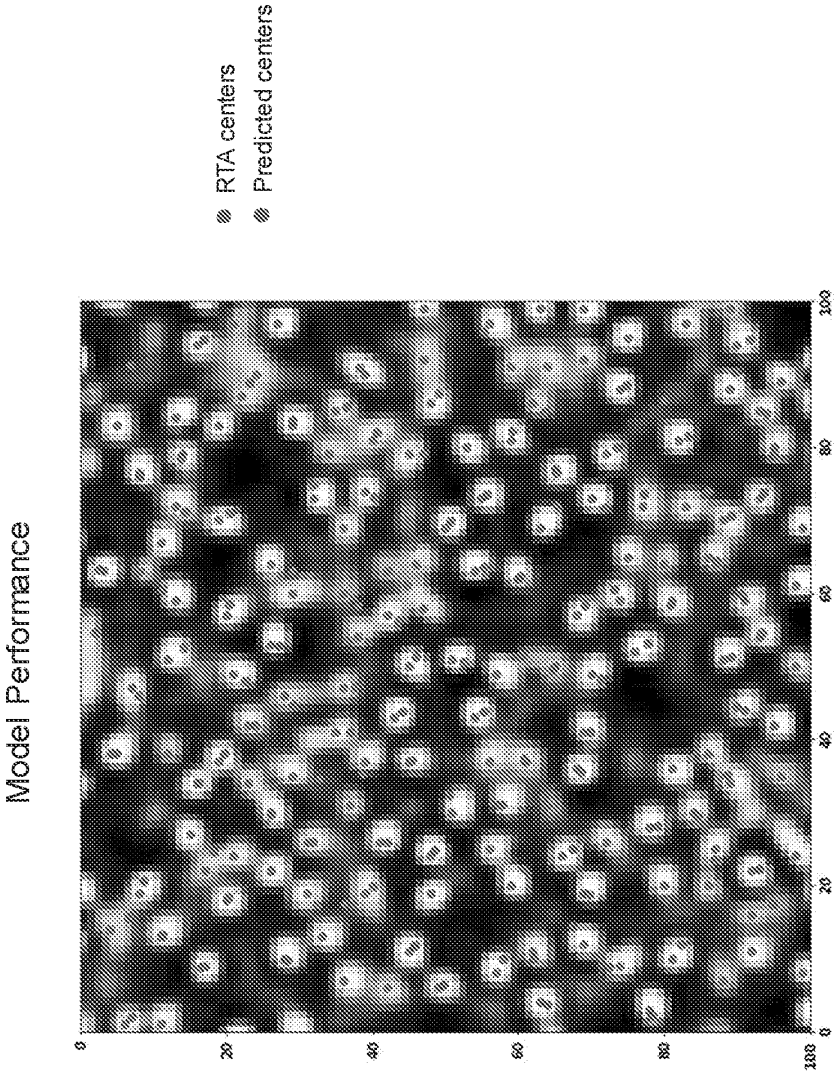


Figure 69

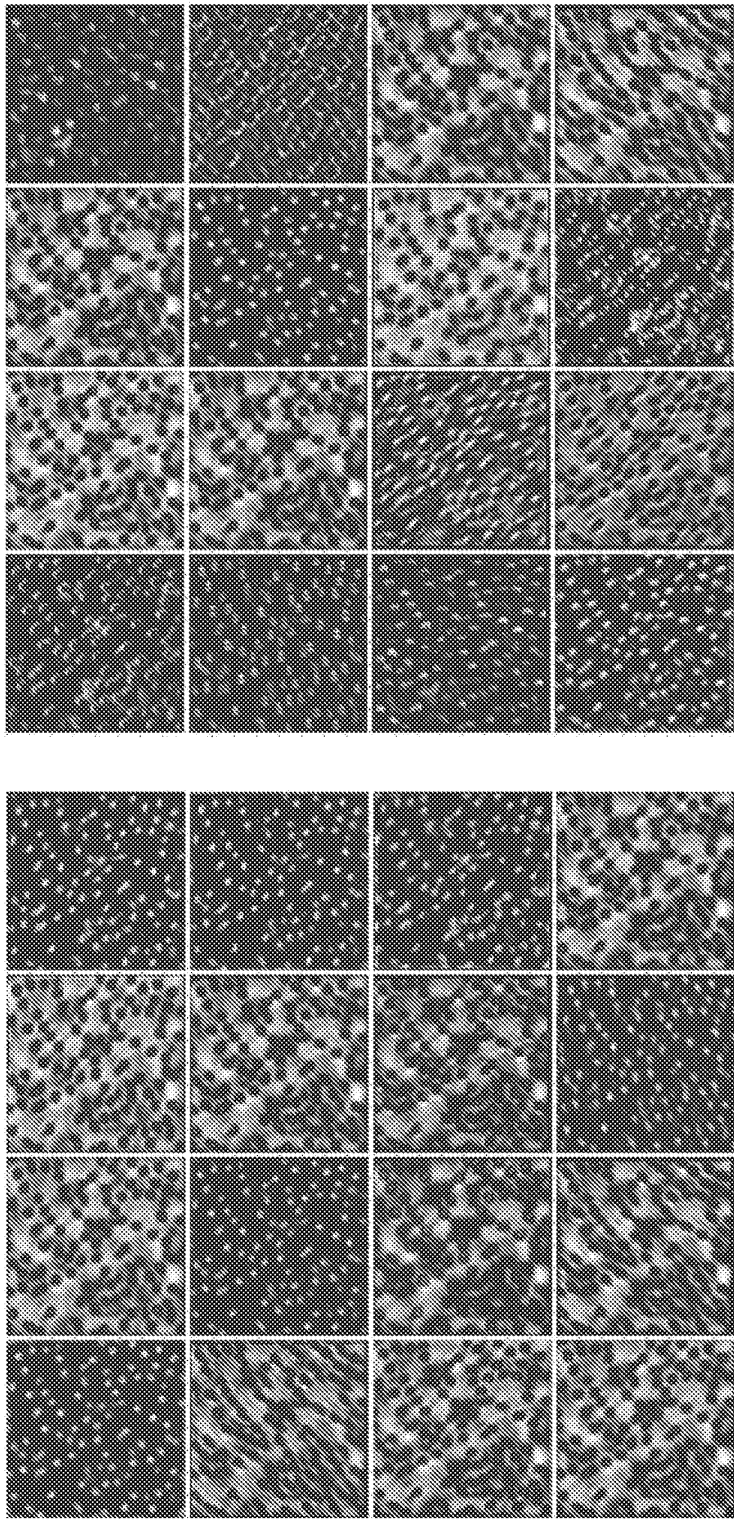
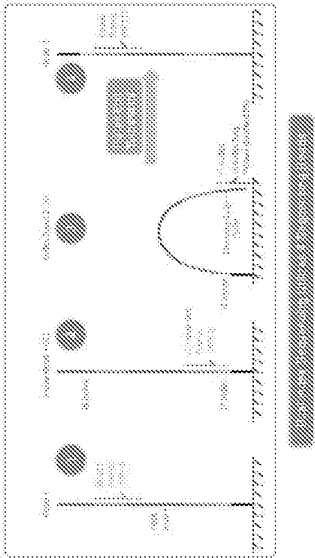


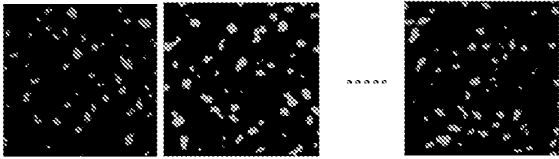
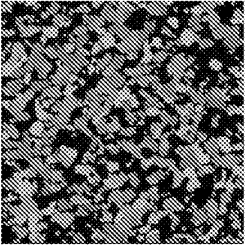
Figure 70

Probe imaging data

Probe imaging using STORM probes



STORM probes – 8 oligos labelled with Alexa-Fluor 647, designed to hybridise to the i7 index region



8 probe images

Datasets

Miseq, Human 40pM
Miseq, Human 80pM
Miseq, E. Coli 40pM

101 cycles of read 1 + i7 index cycles

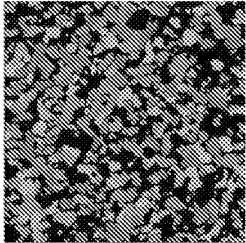


Figure 71

Using beads for image registration

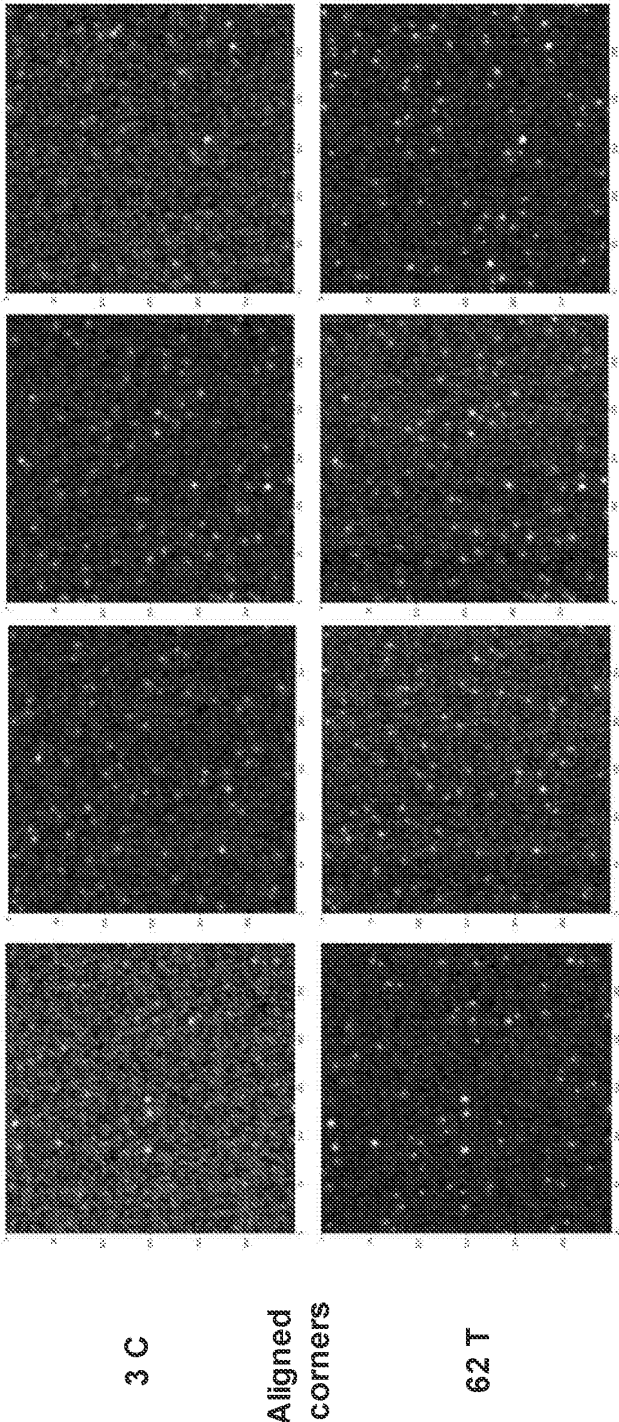


Figure 72

Cluster Statistics

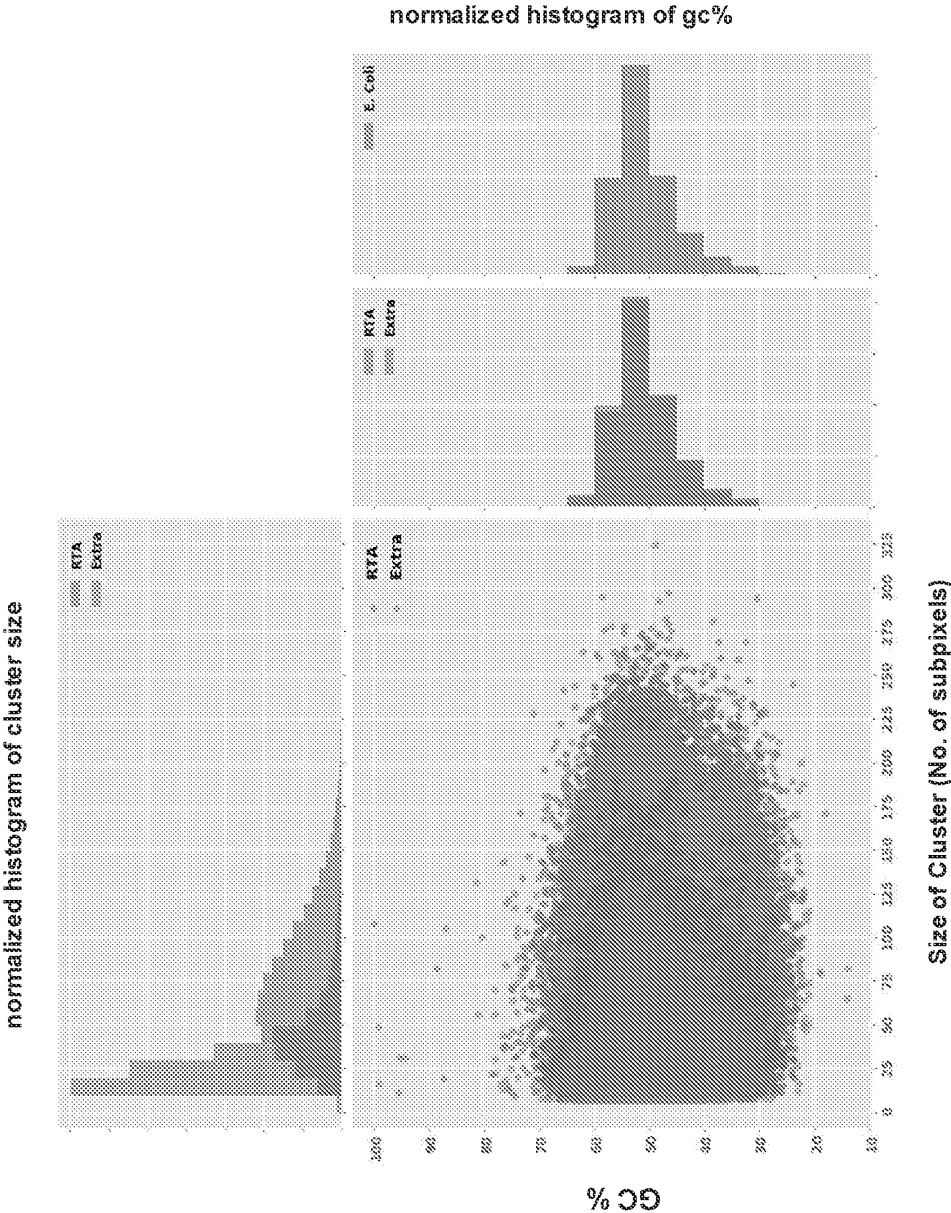
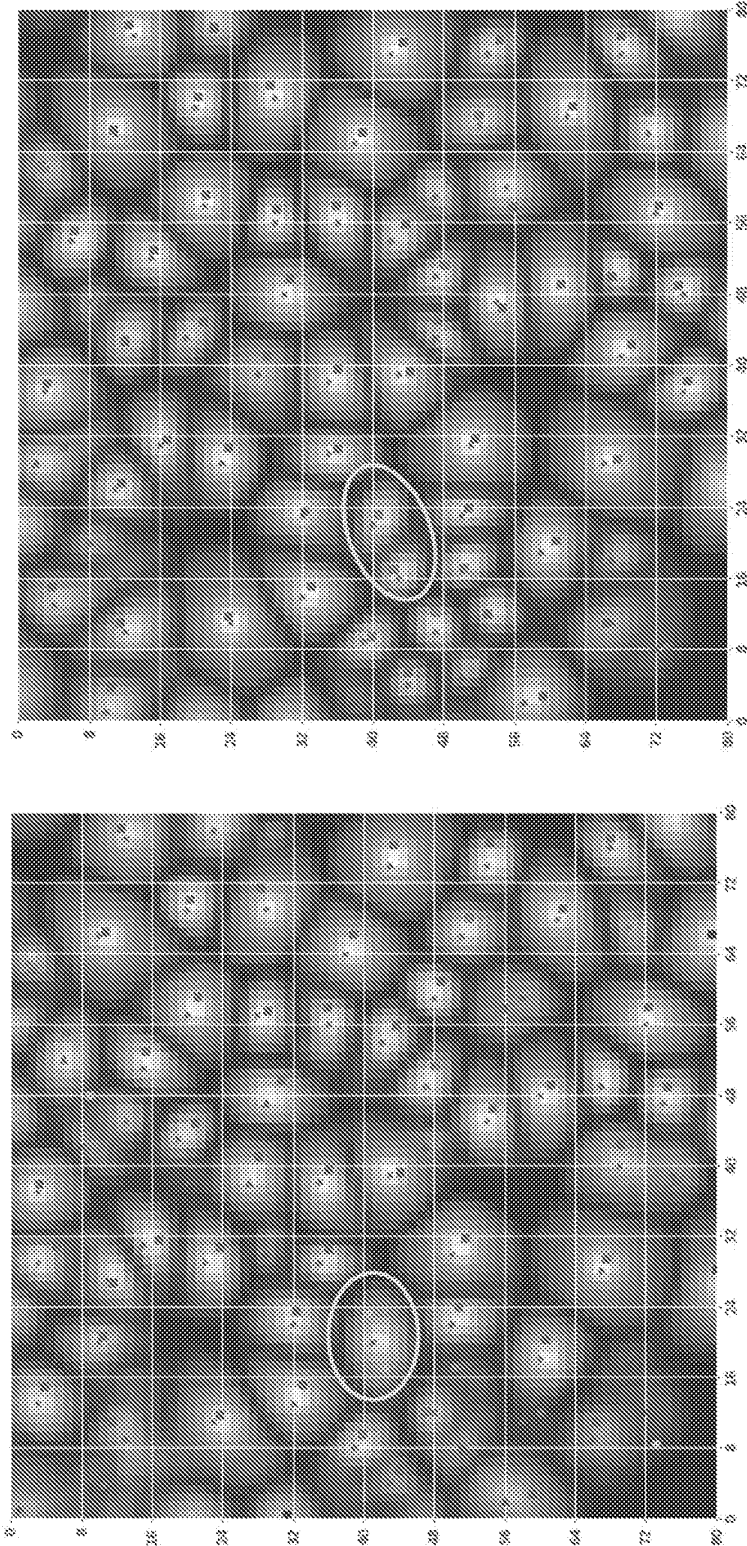


Figure 73

Model Predictions



7 cycles: two clusters

5 cycles: single cluster;
same sequence

Figure 74

RTA centers vs. mass centers

RTA centers

956054 + 0 in total (QC-passed reads + QC-failed reads)
164 + 0 secondary
0 + 0 supplementary
1045 + 0 duplicates
932910 + 0 mapped (97.58% : N/A)
955890 + 0 paired in sequencing
477945 + 0 read1
477945 + 0 read2
923322 + 0 properly paired (96.59% : N/A)
925276 + 0 with itself and mate mapped
7470 + 0 singletons (0.78% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

bases mapped: 140844646 # ignores clipping
bases mapped (cigar): 139683478
bases trimmed: 0
bases duplicated: 157795
mismatches: 821045 # from NM fields
error rate: 5.877897e-03 # mismatches / bases mapped (cigar)

Mass centers

956048 + 0 in total (QC-passed reads + QC-failed reads)
158 + 0 secondary
0 + 0 supplementary
841 + 0 duplicates
936609 + 0 mapped (97.97% : N/A)
955890 + 0 paired in sequencing
477945 + 0 read1
477945 + 0 read2
928926 + 0 properly paired (97.18% : N/A)
930732 + 0 with itself and mate mapped
5719 + 0 singletons (0.60% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

bases mapped: 141404101 # ignores clipping
bases mapped (cigar): 140673703
bases trimmed: 0
bases duplicated: 126991
mismatches: 498799 # from NM fields
error rate: 3.545787e-03 # mismatches / bases mapped (cigar)

Figure 75

Extra grown clusters

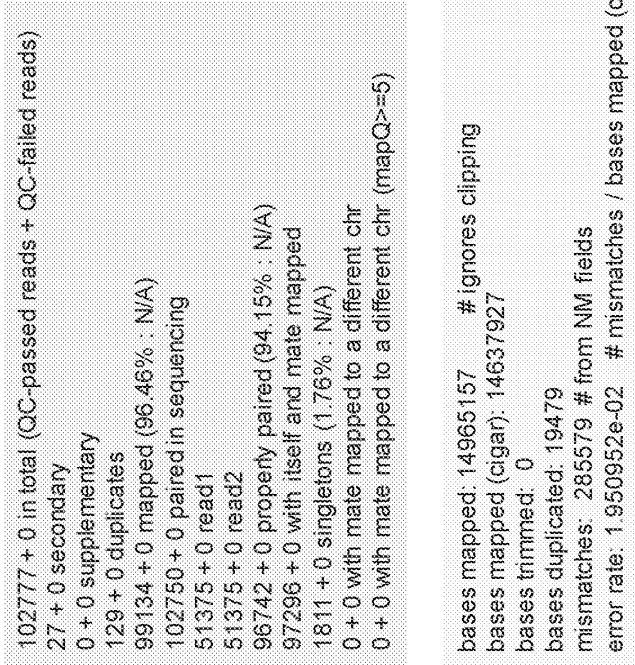


Figure 76

Data sets

Nextseq

Library	RTA PF%
1.3pM	61%
1.3pM	42%
3pM	14%

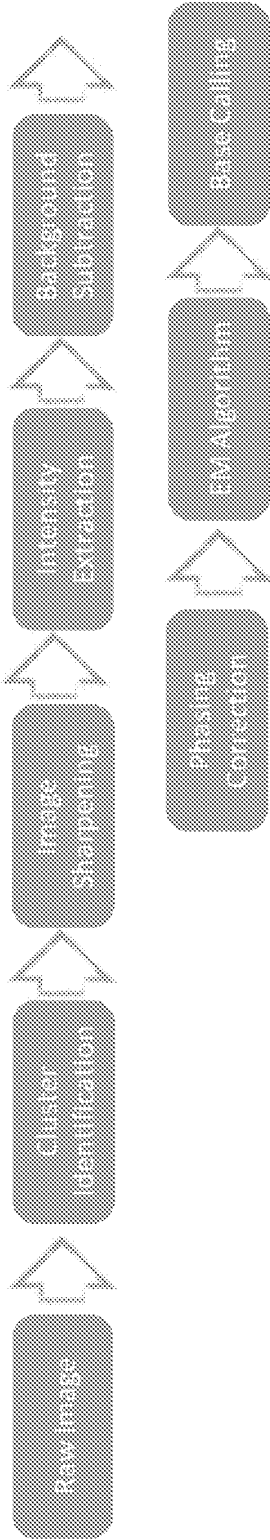
Miseq

Library	Density	RTA PF%
20pM	1254 K/mm2	85%
30pM	1524 K/mm2	72%
40pM	1529 K/mm2	31%

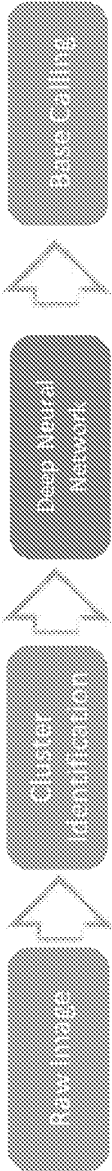
Figure 77

Base Calling

RTA



Deep Learning



Deep Learning

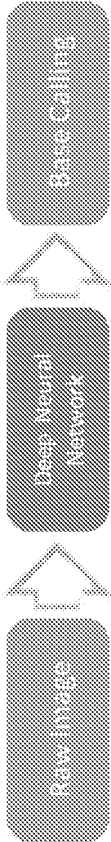


Figure 78

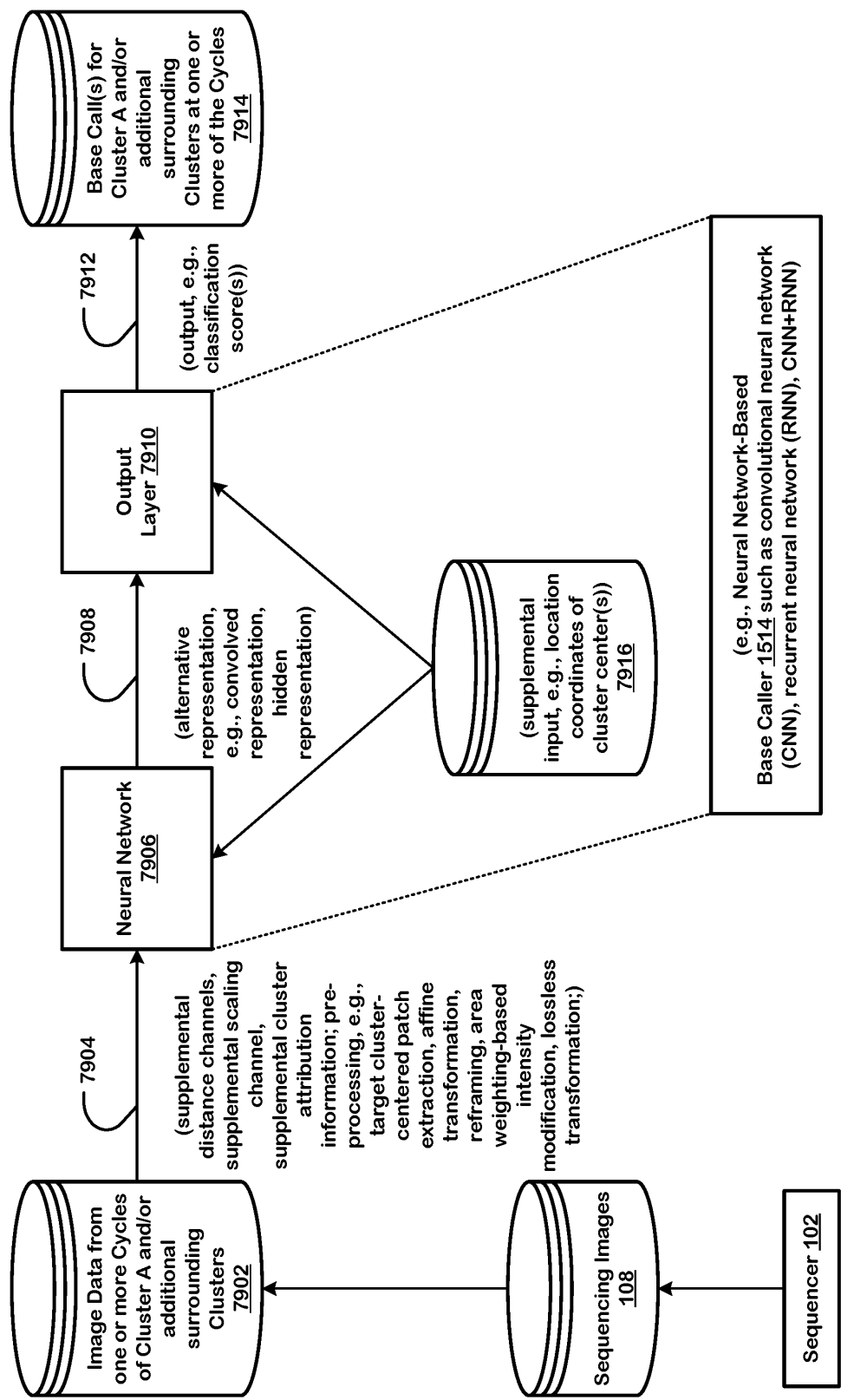


Figure 79

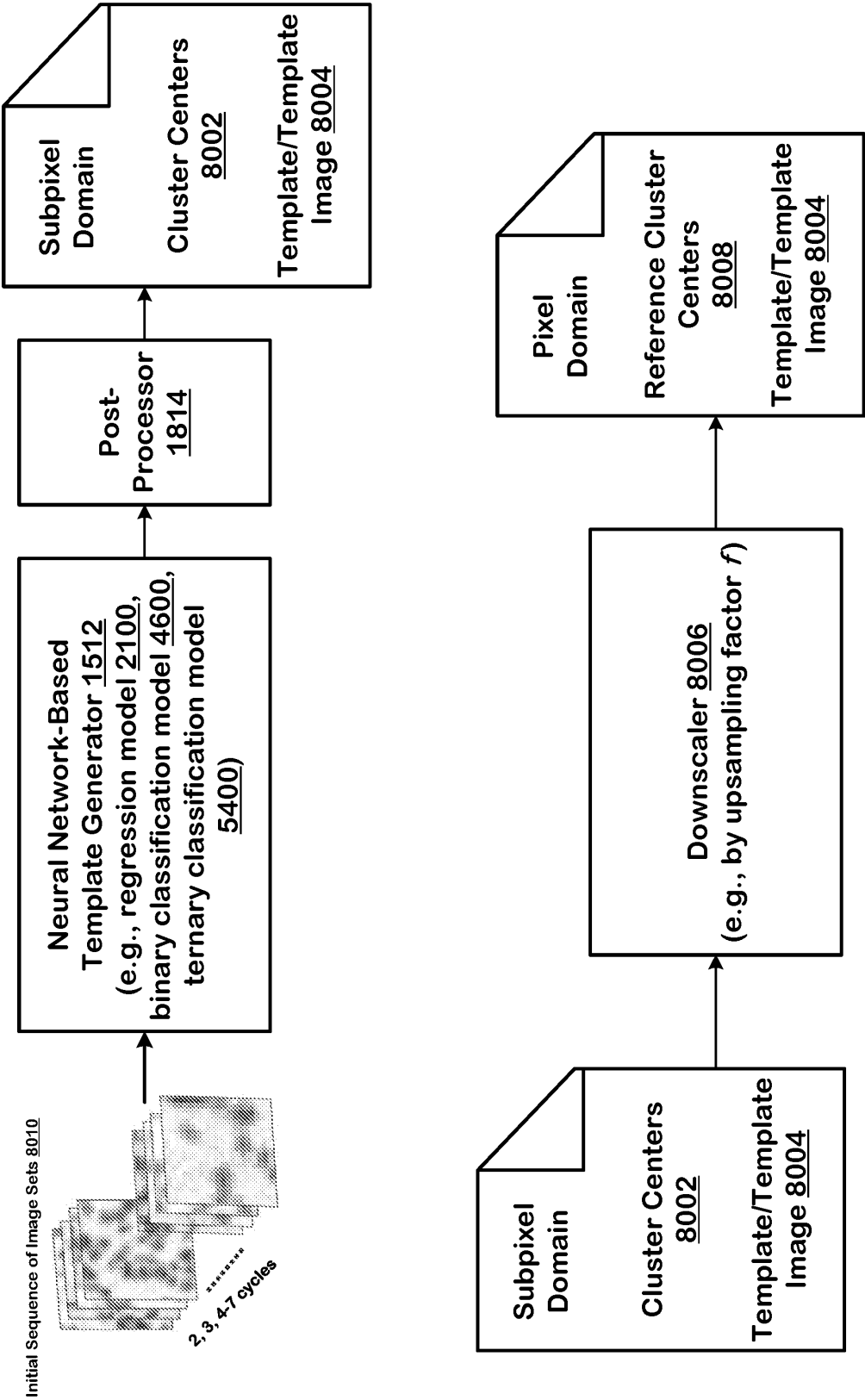


Figure 80

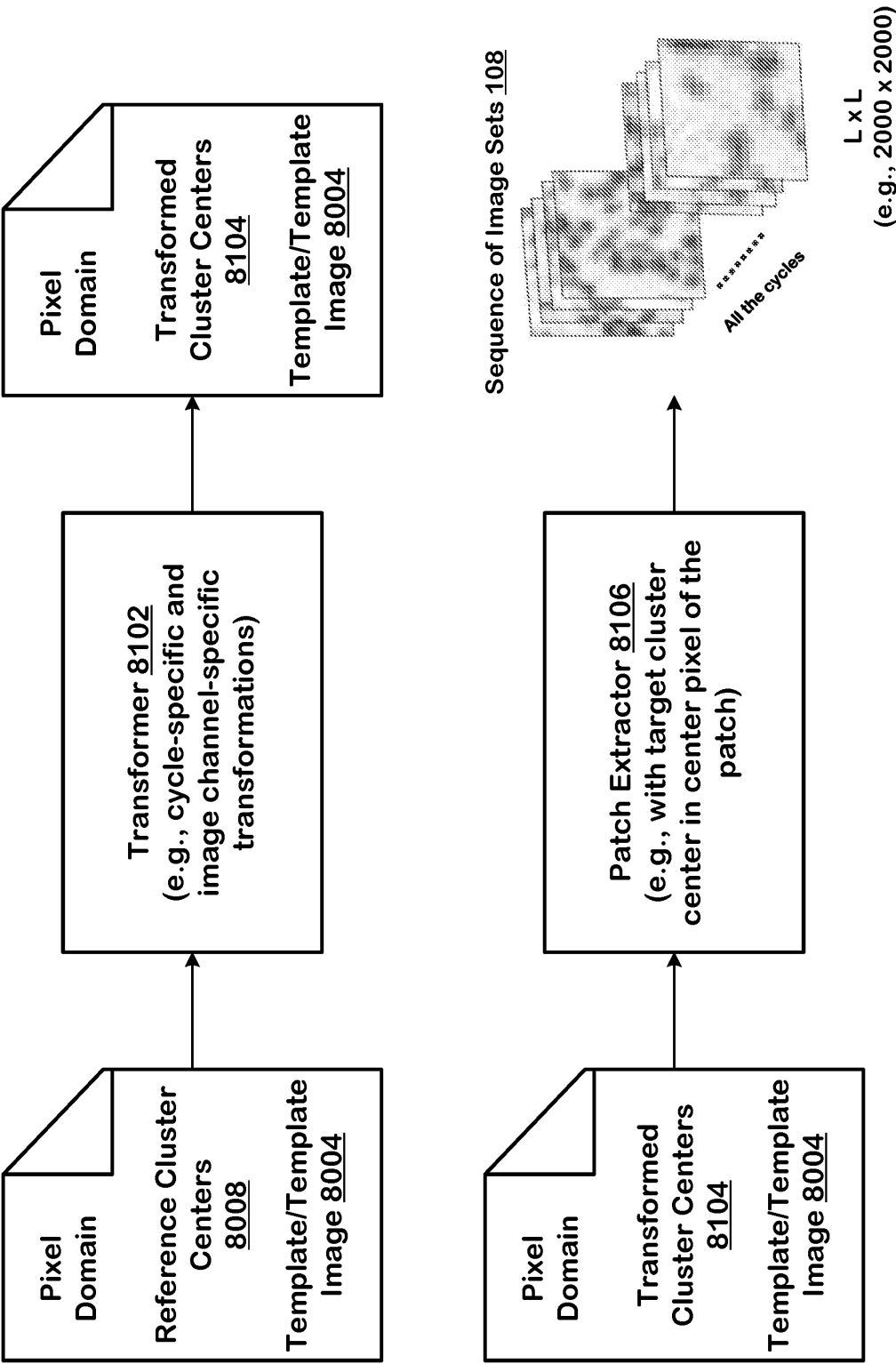


Figure 81

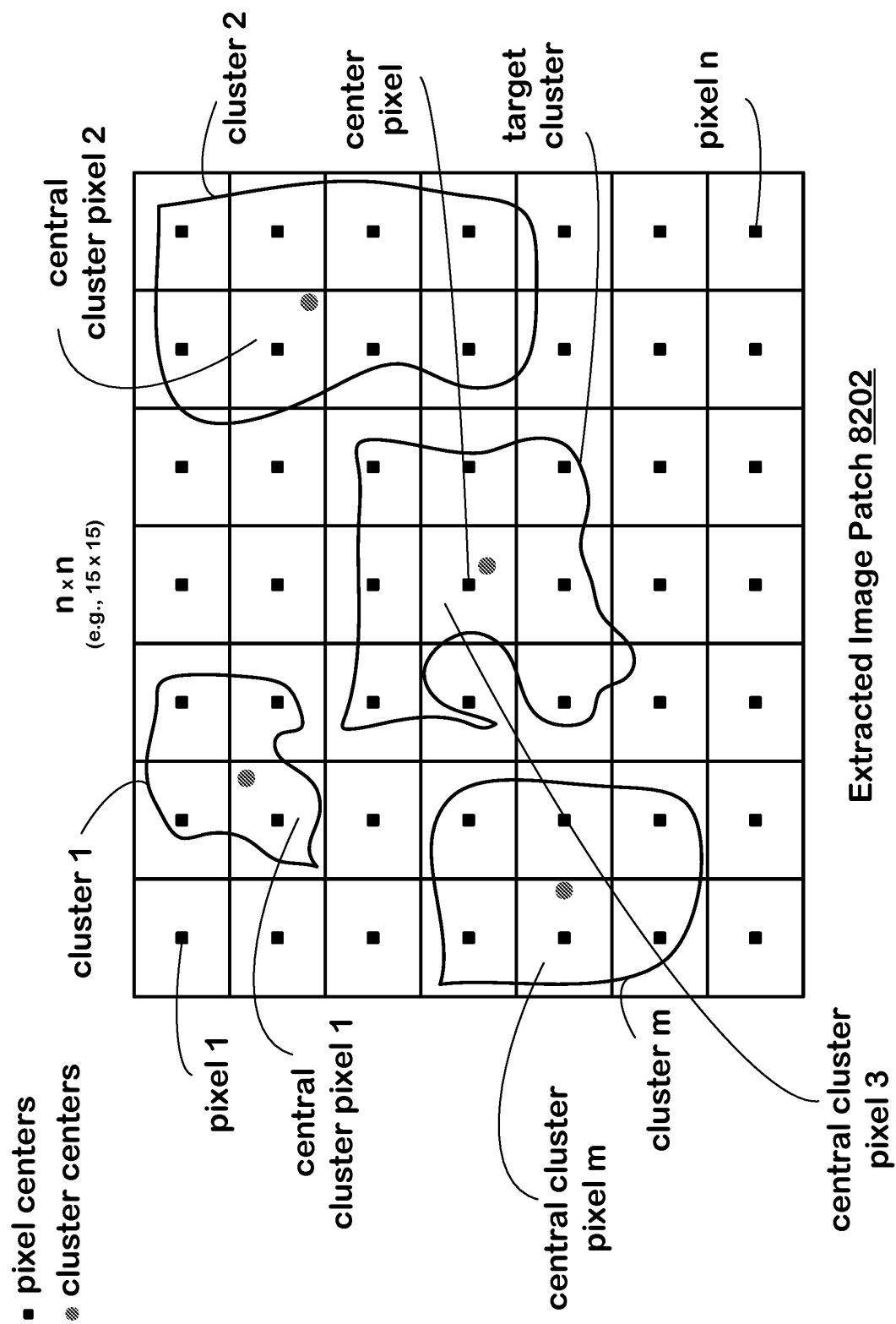


Figure 82

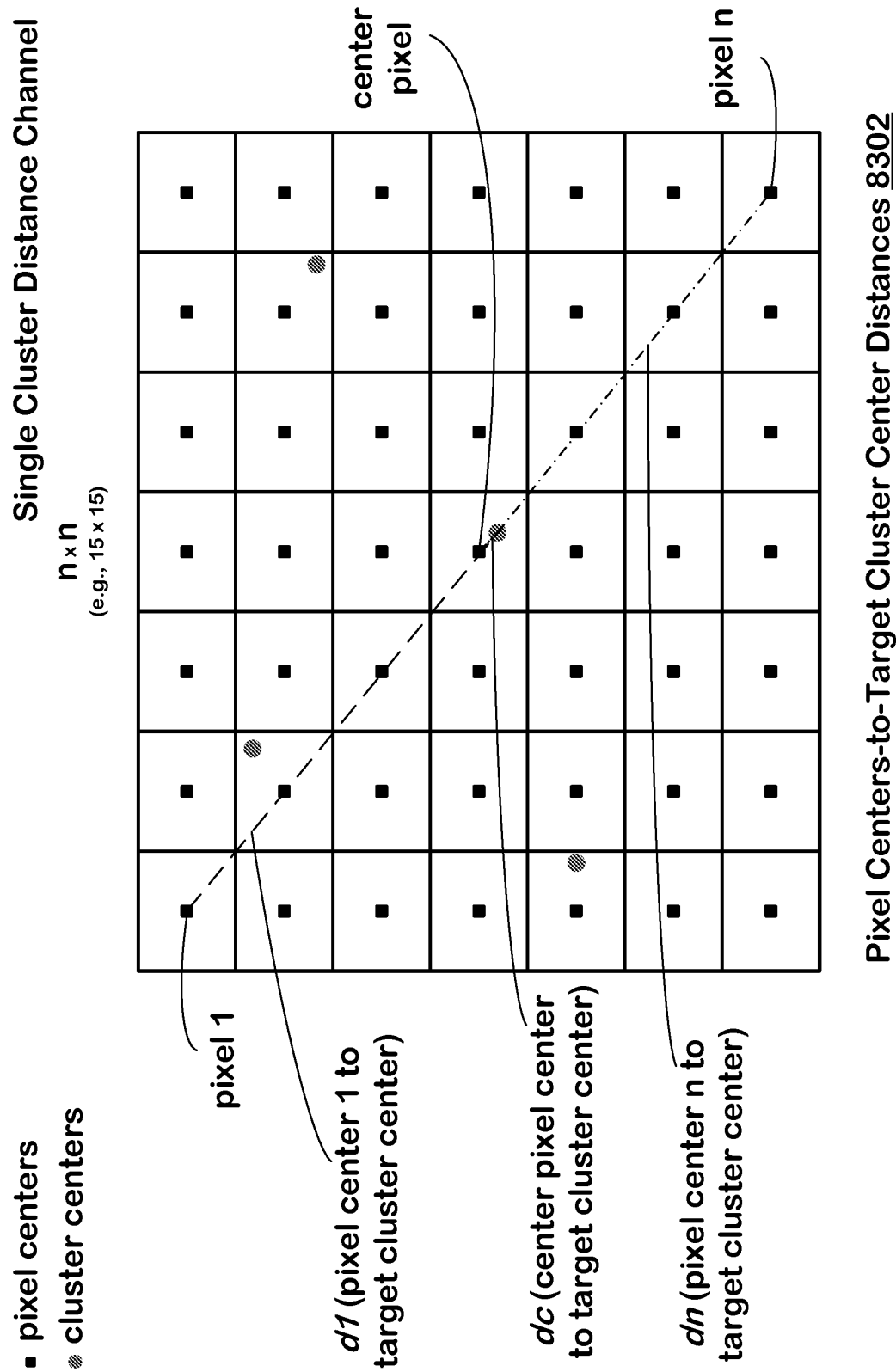


Figure 83

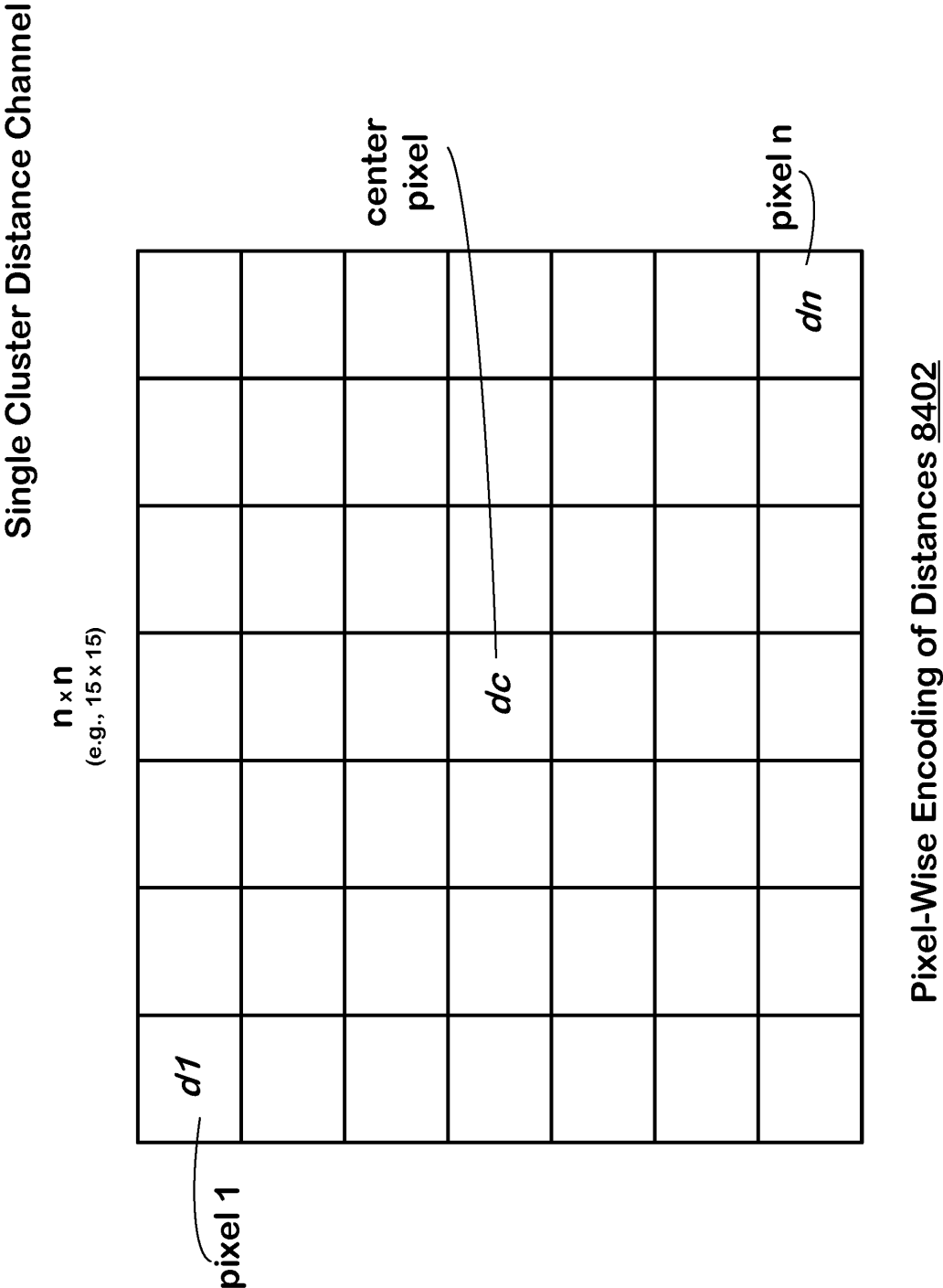
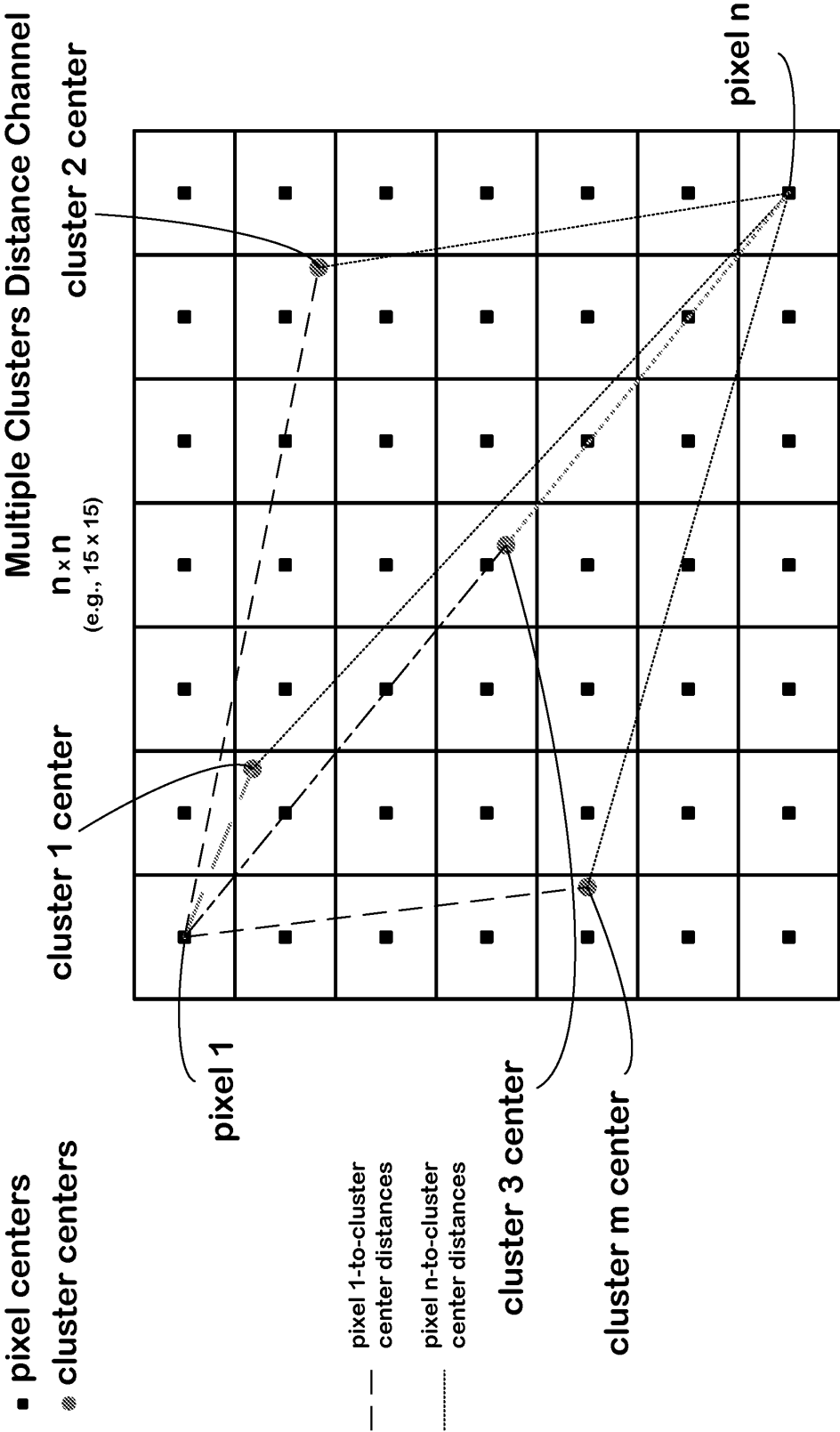
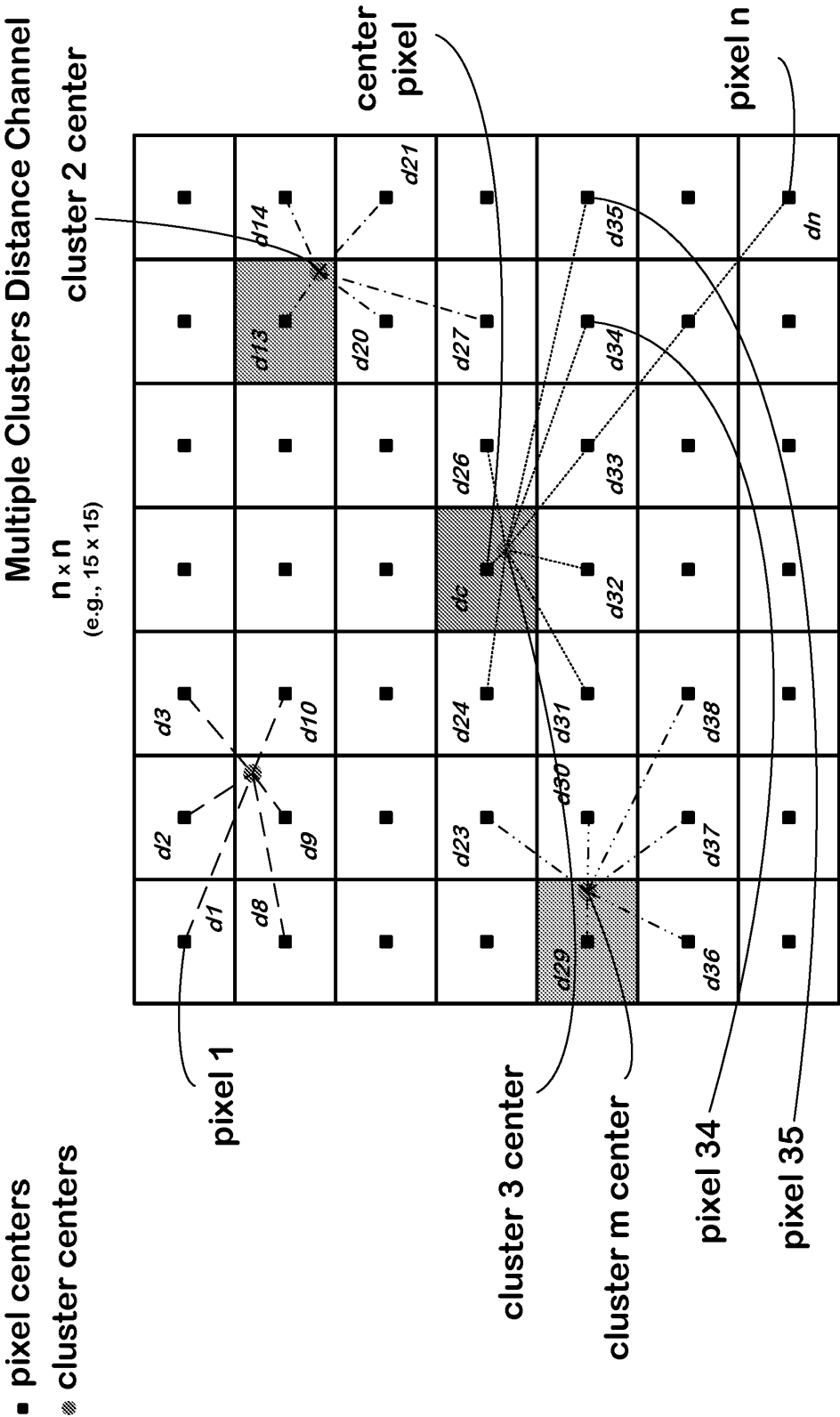


Figure 84



Pixel Center-to-Cluster Center Distances 8502

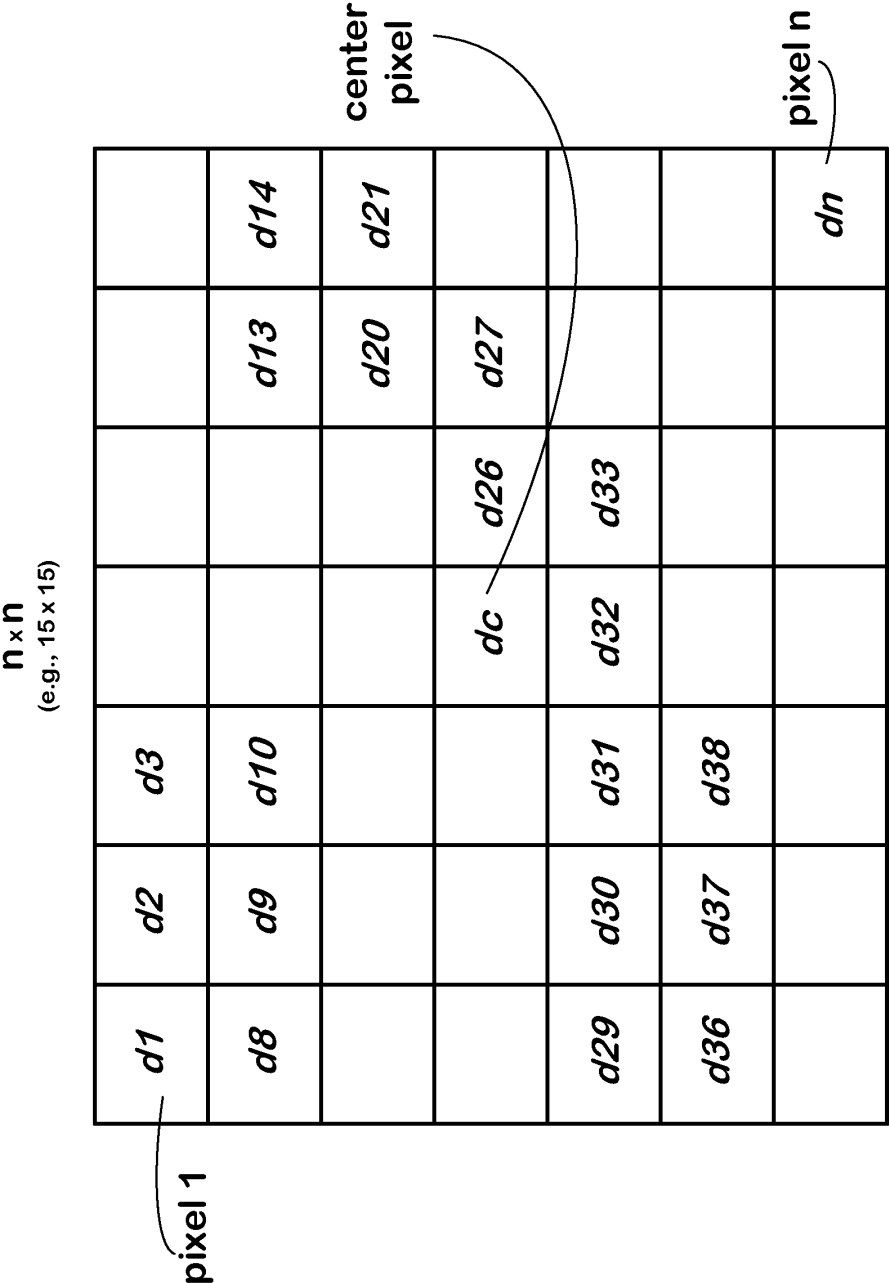
Figure 85a



Pixel Center-to-Nearest Cluster Center Distances 8504

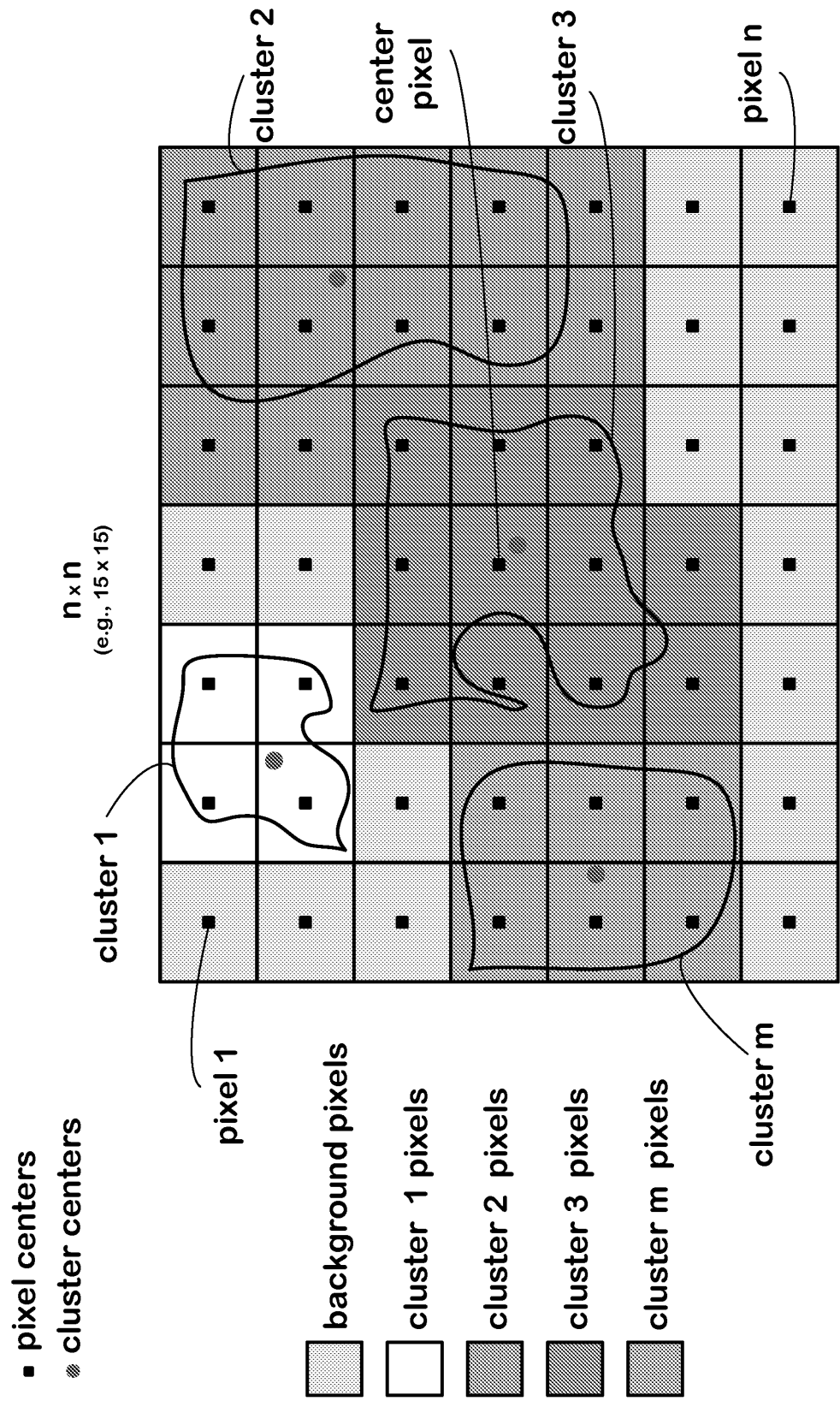
Figure 85b

Multiple Clusters Distance Channel



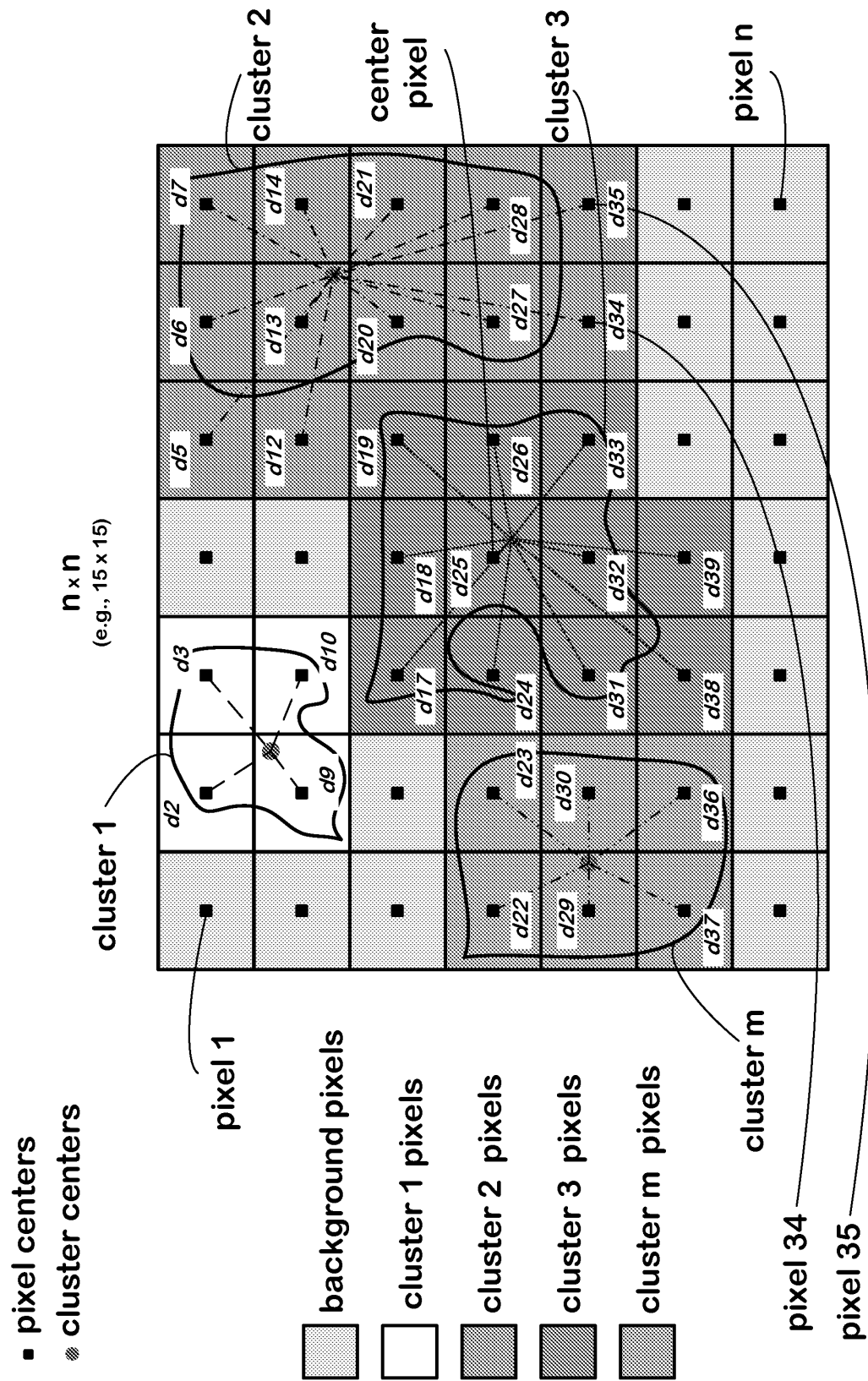
Pixel-Wise Encoding of Distances 8602

Figure 86



Pixel-to-Cluster Classification 8702

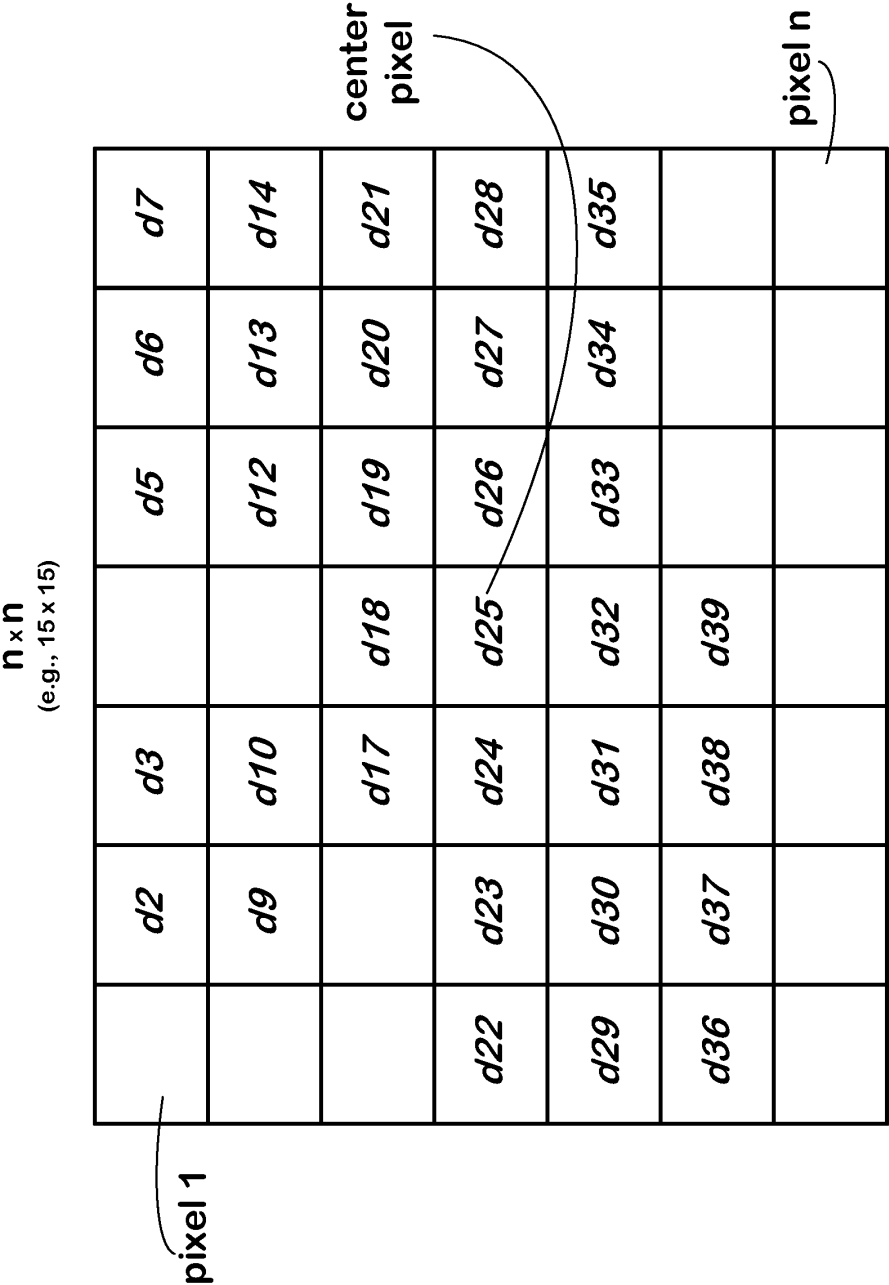
Figure 87



Pixel Centers-to-Assigned Cluster Center Distances 8802

Figure 88

Multiple Clusters, Shape-Based Distance Channel



Pixel-Wise Encoding of Distances 8902

Figure 89

Specialized Architecture

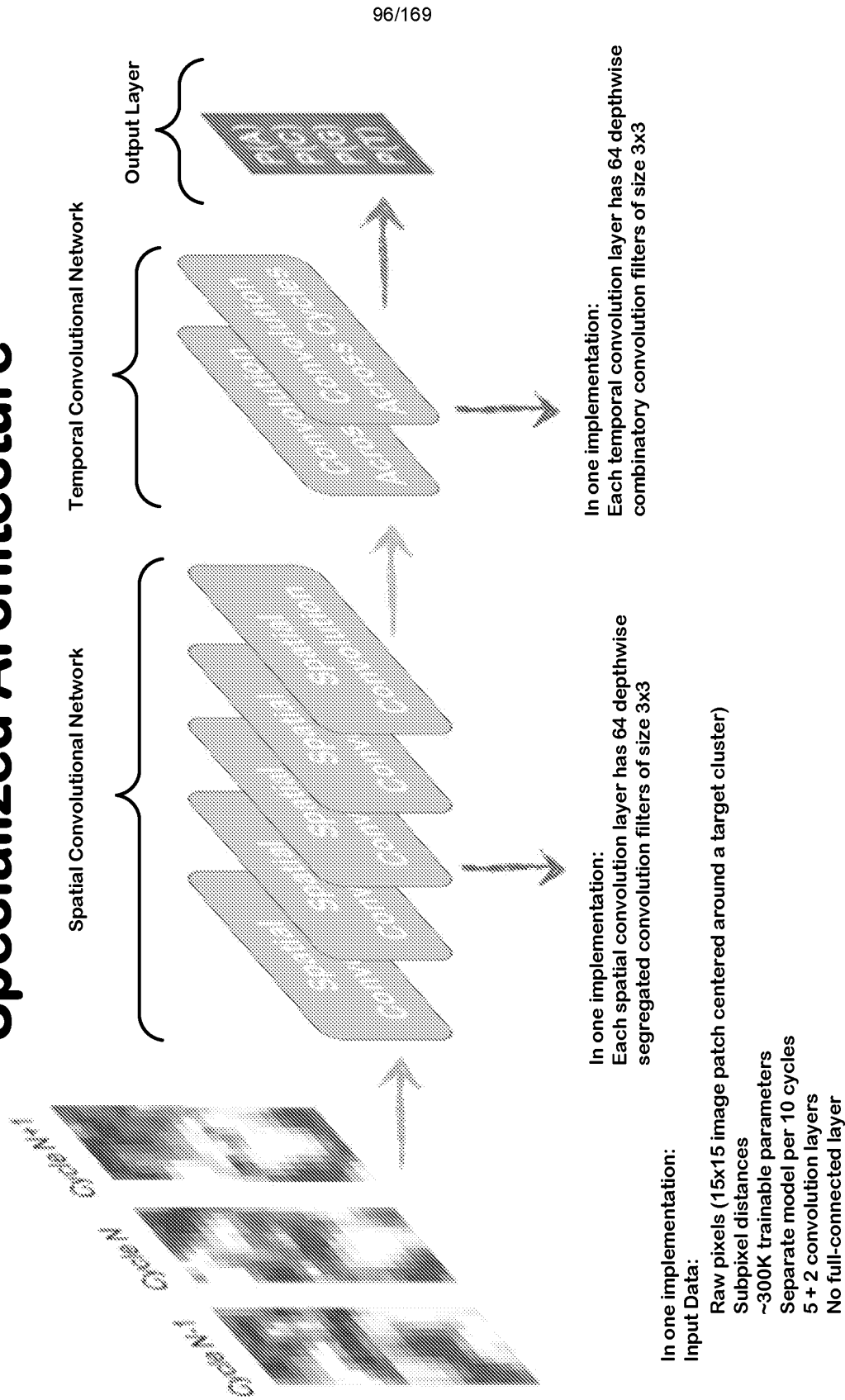


Figure 90

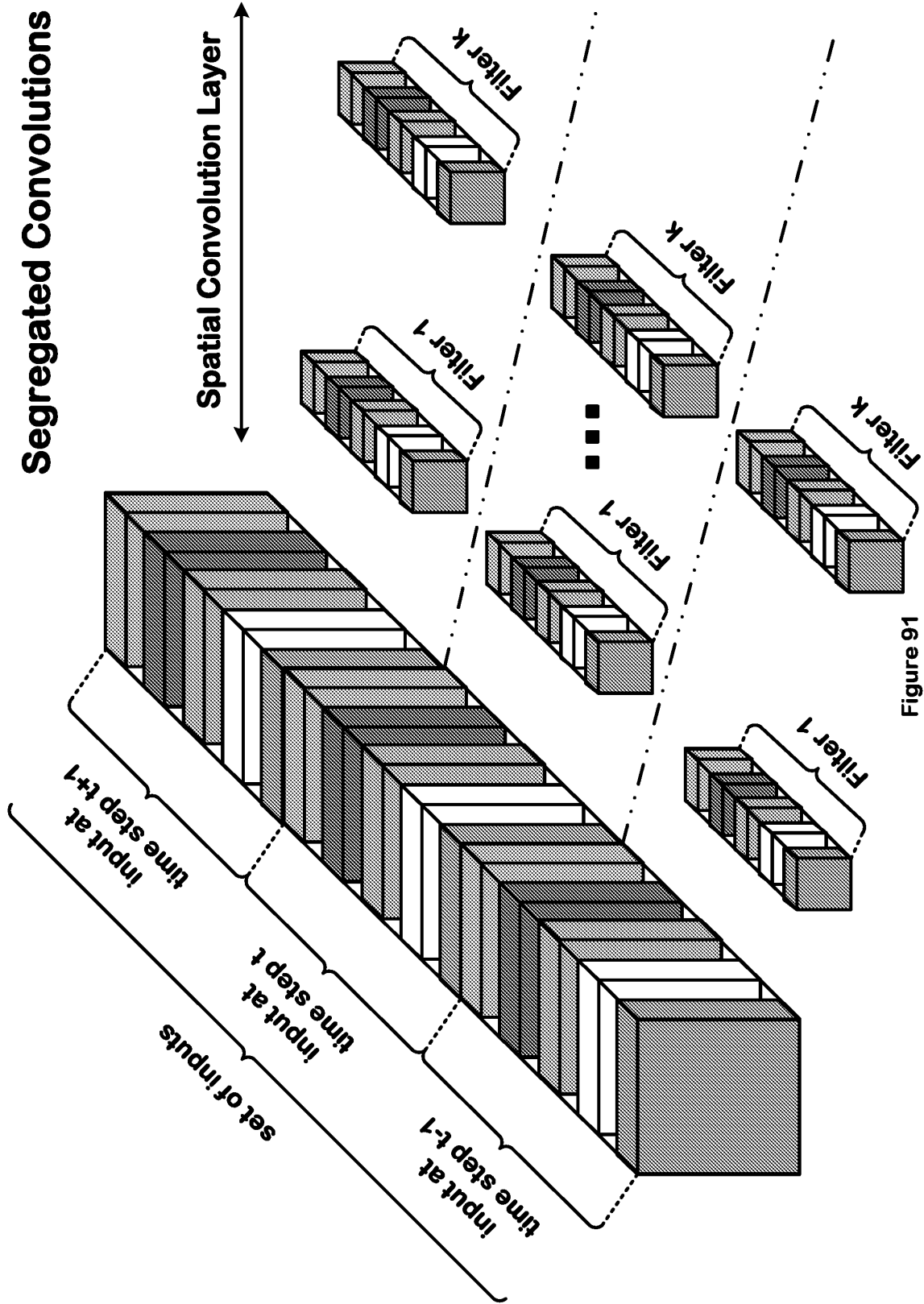
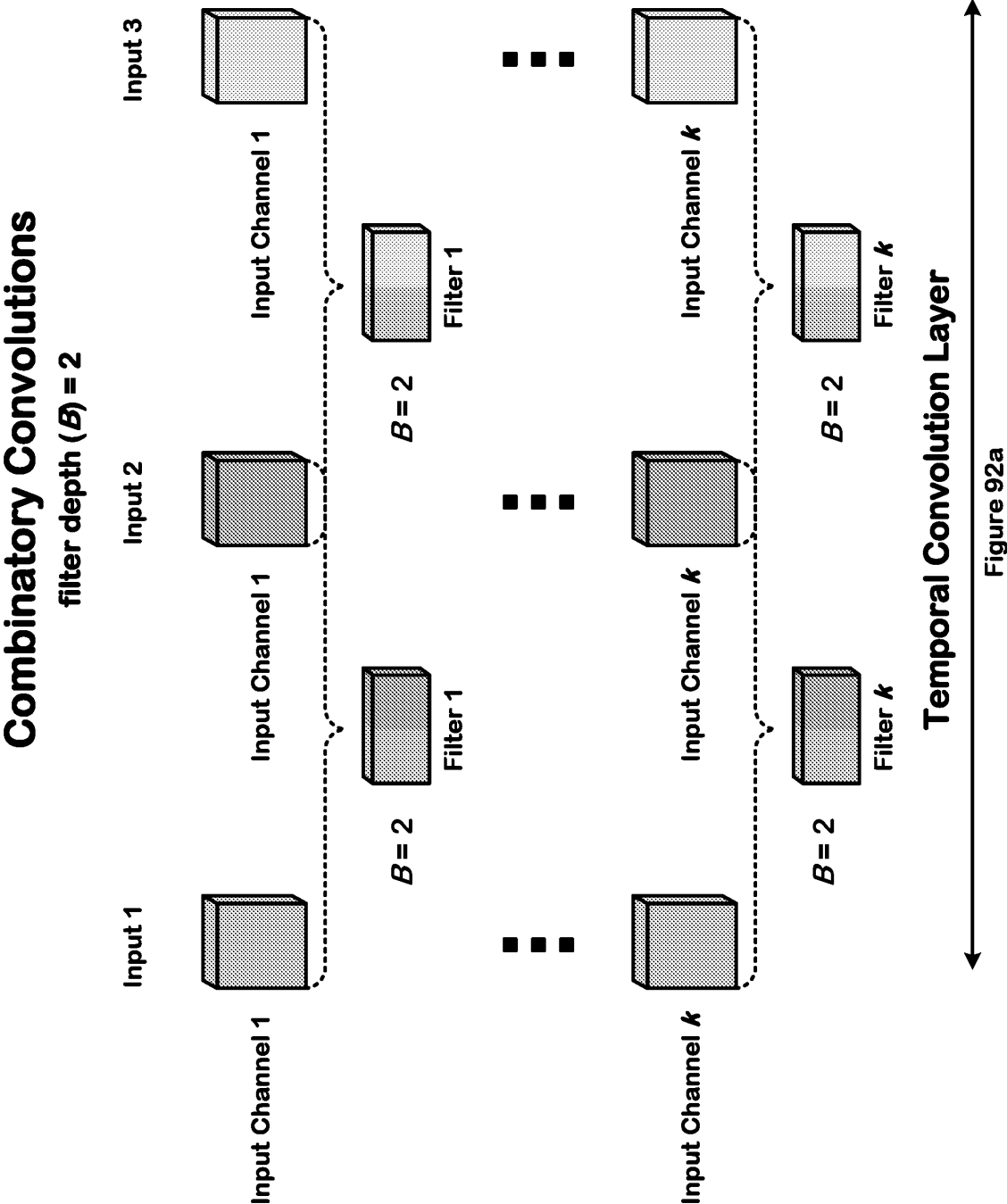


Figure 91



$B = 2$

Filter 1

$B = 2$

Filter 1

$B = 2$

Filter 1

Input Channel k

Input Channel k

Input Channel k

$B = 2$

Filter k

$B = 2$

Filter k

$B = 2$

Filter k

$B = 2$

Filter k

$B = 2$

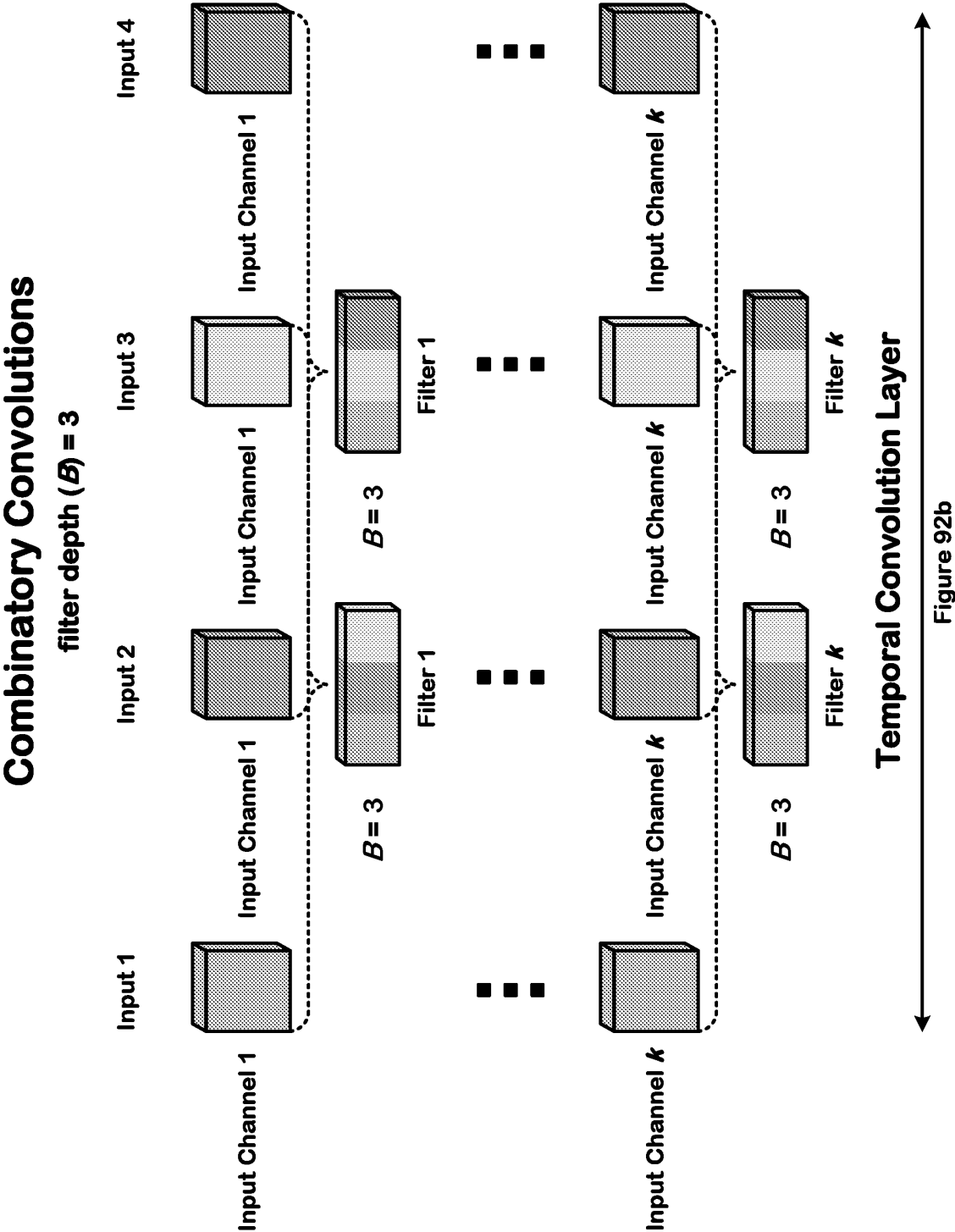
Filter k

$B = 2$

Filter k

Temporal Convolution Layer

Figure 92a



Temporal Convolution Layer

Figure 92b

Filter Banks

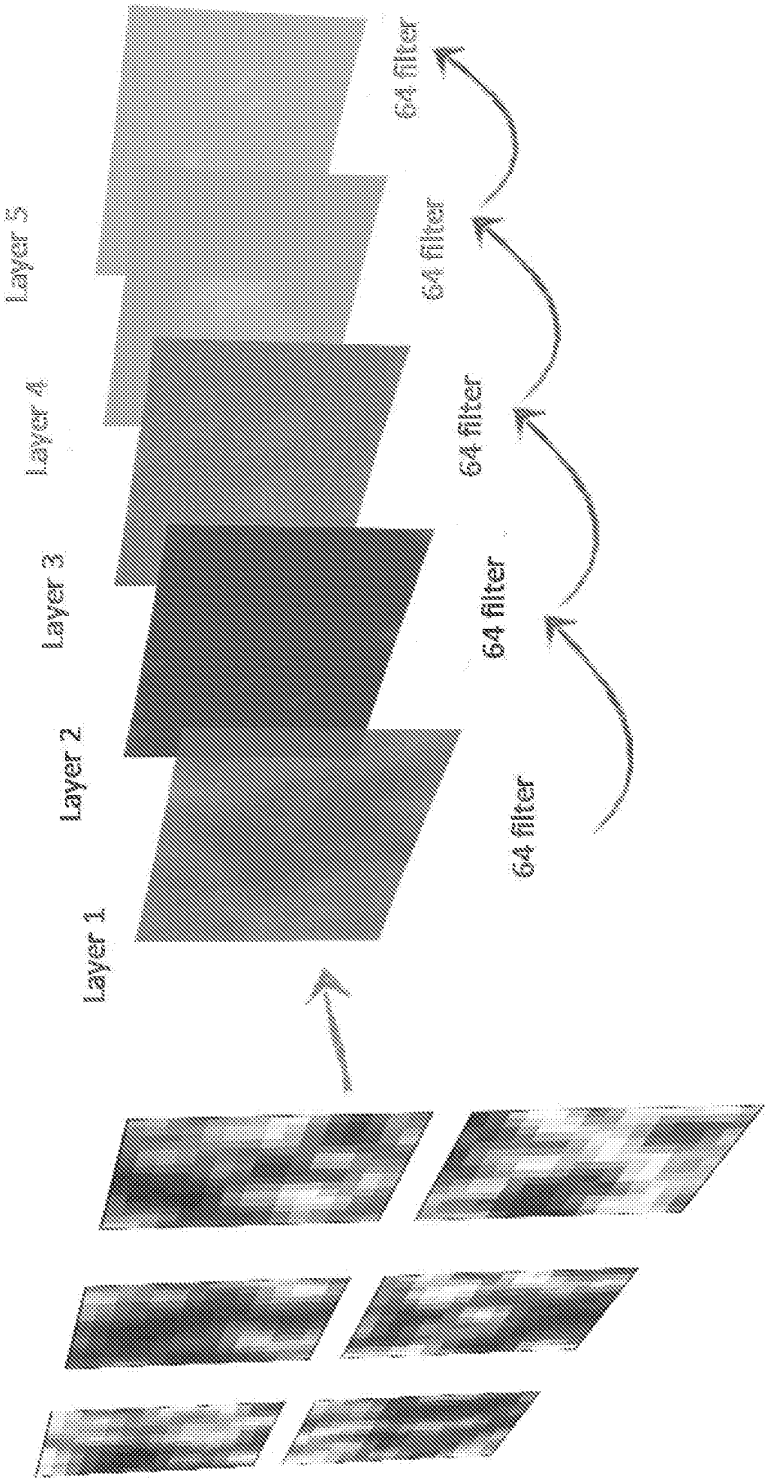


Figure 93

Scaling Channel Calculation

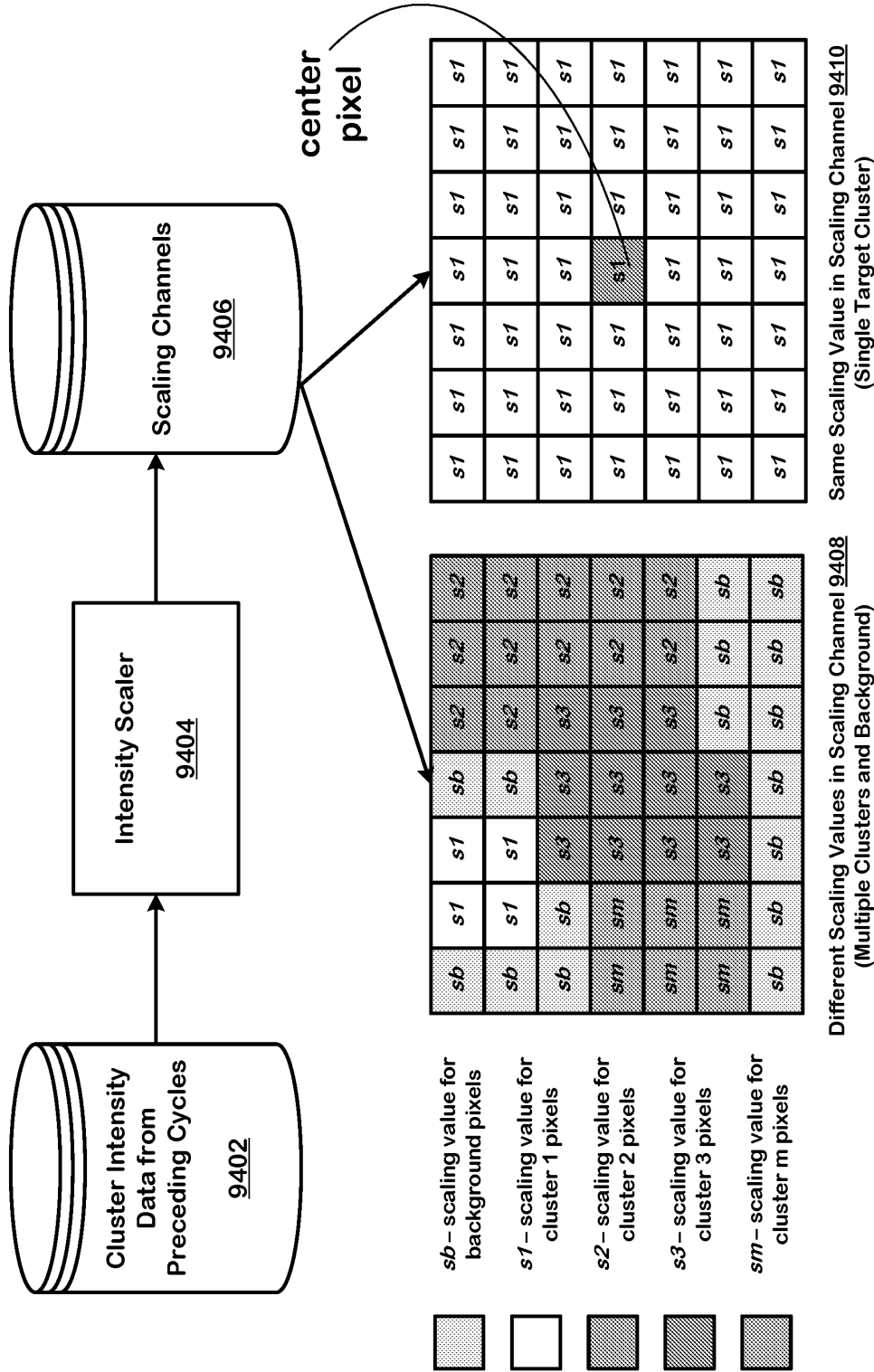


Figure 94

Input Data for a Single Sequencing Cycle that produces a Red Image and a Green Image 9500

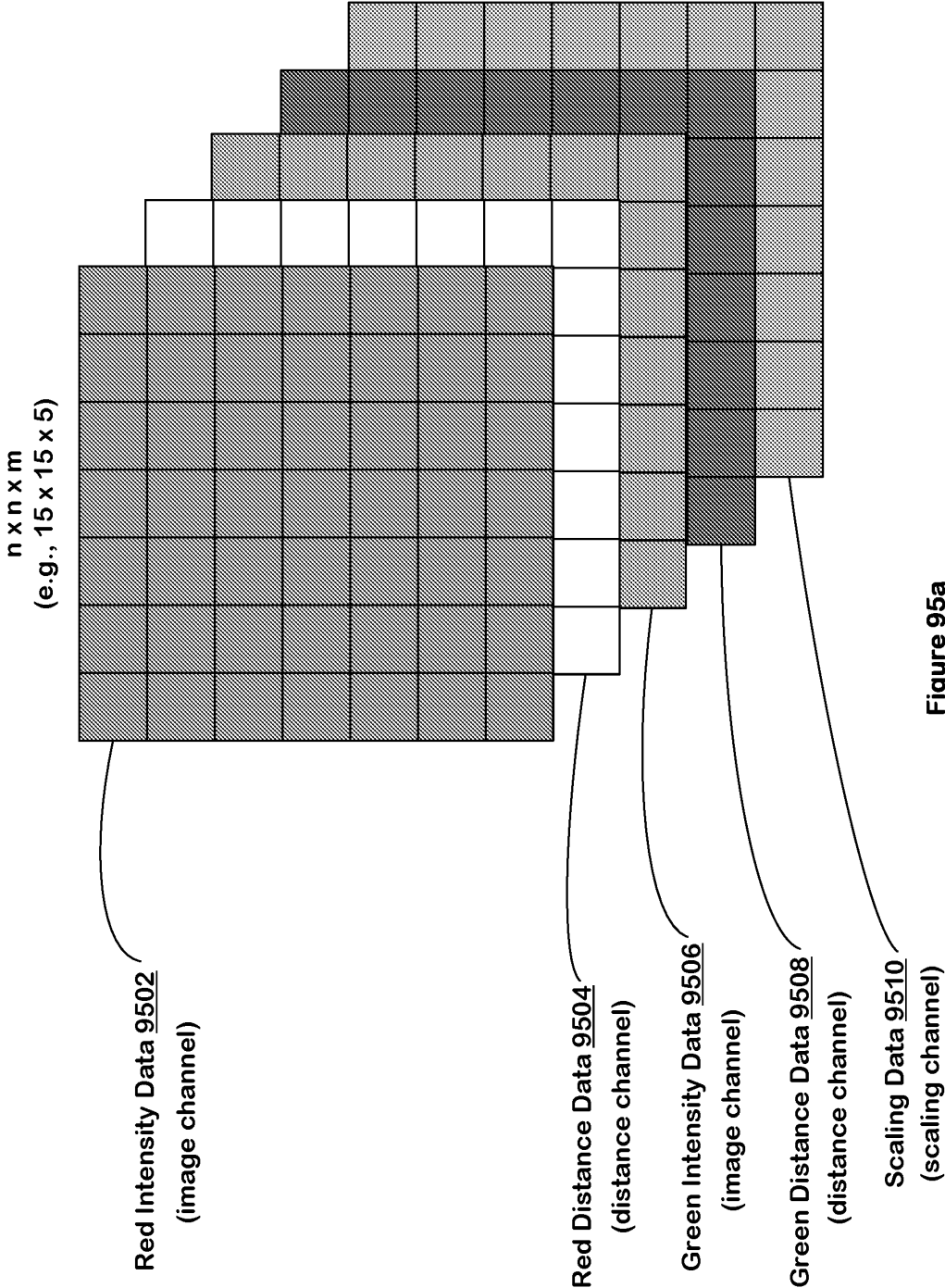
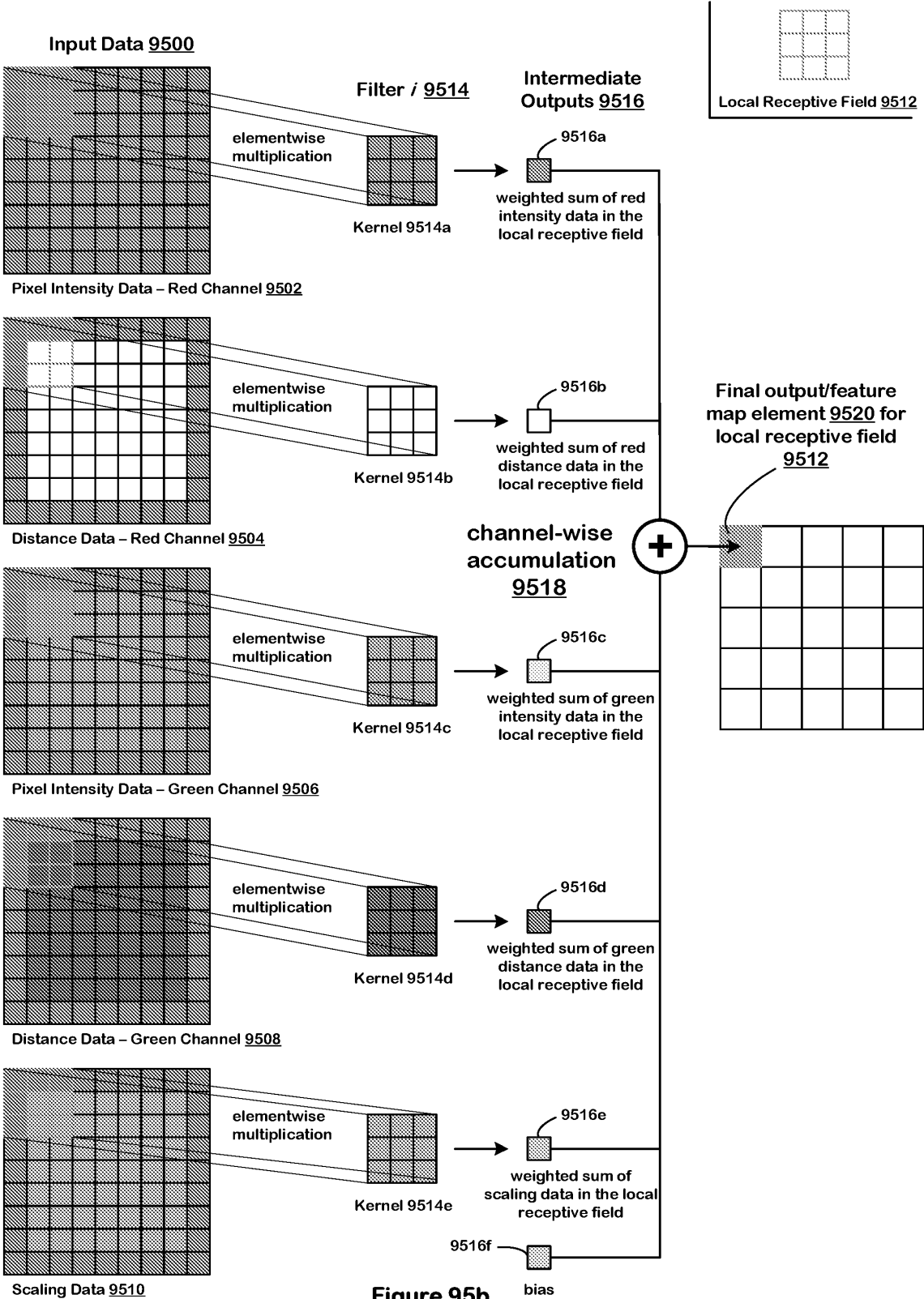


Figure 95a



Single Cluster Base Calling – Spatial Convolution Layers

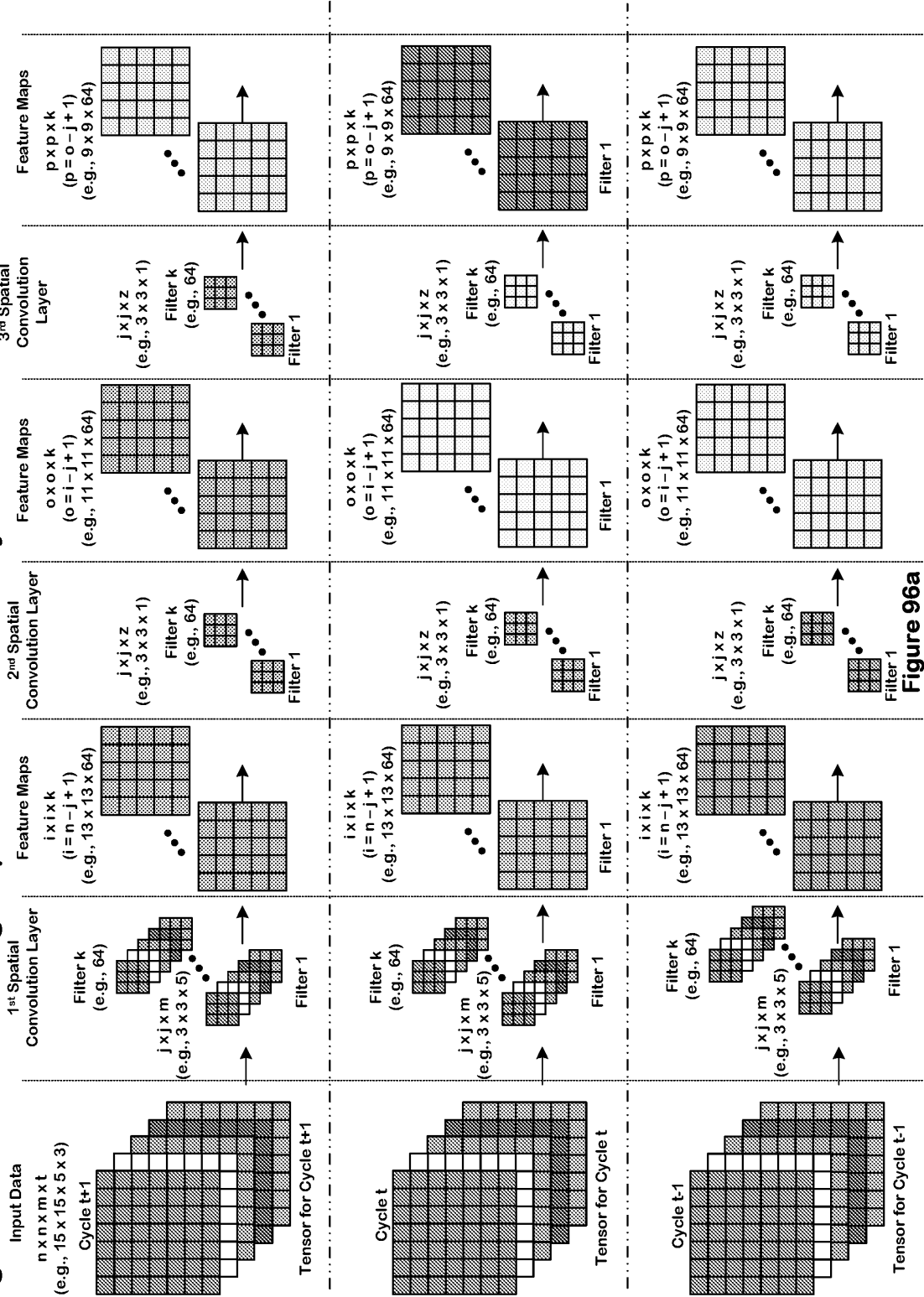


Figure 96a

Single Cluster Base Calling – Spatial Convolution Layers

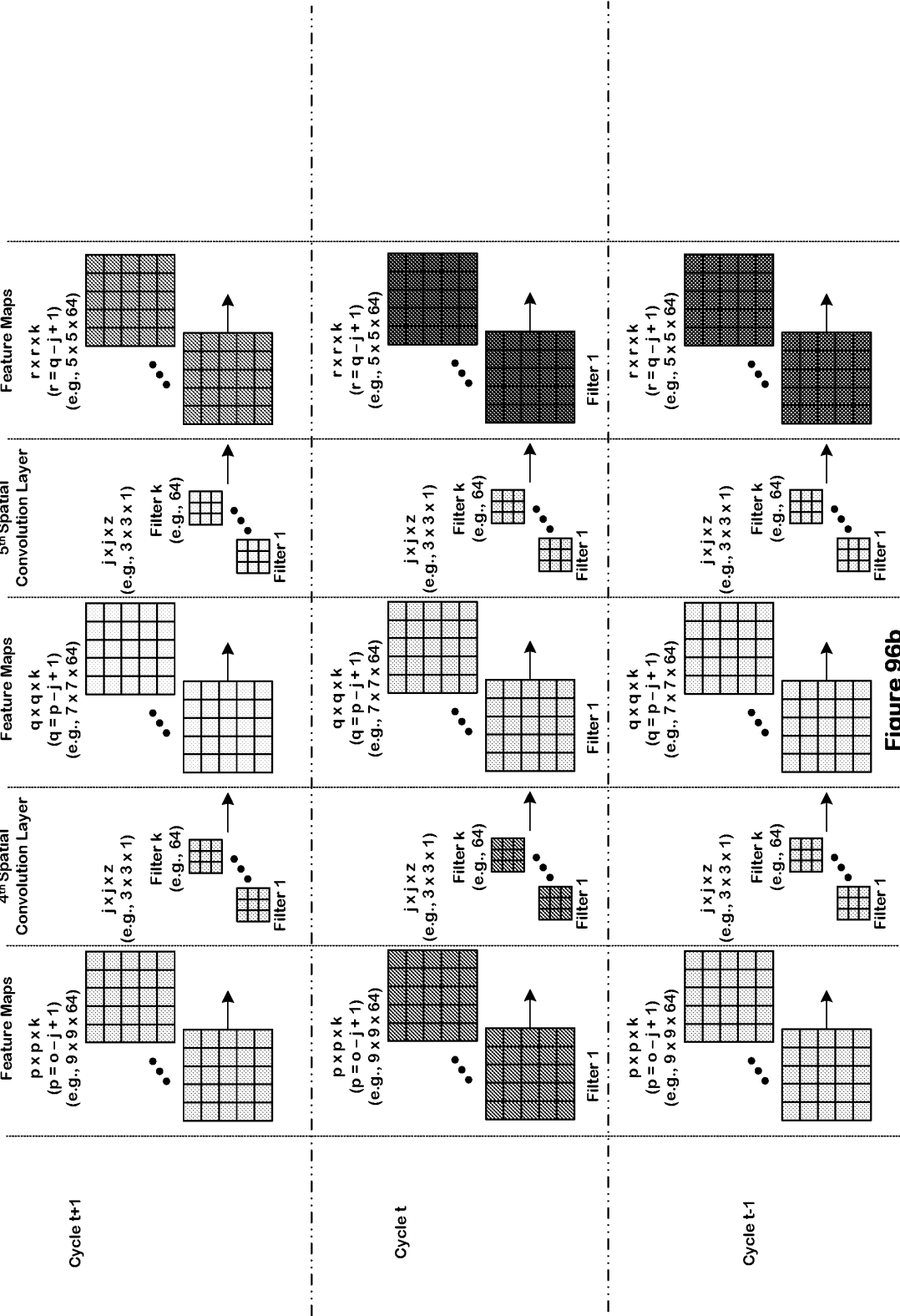


Figure 96b

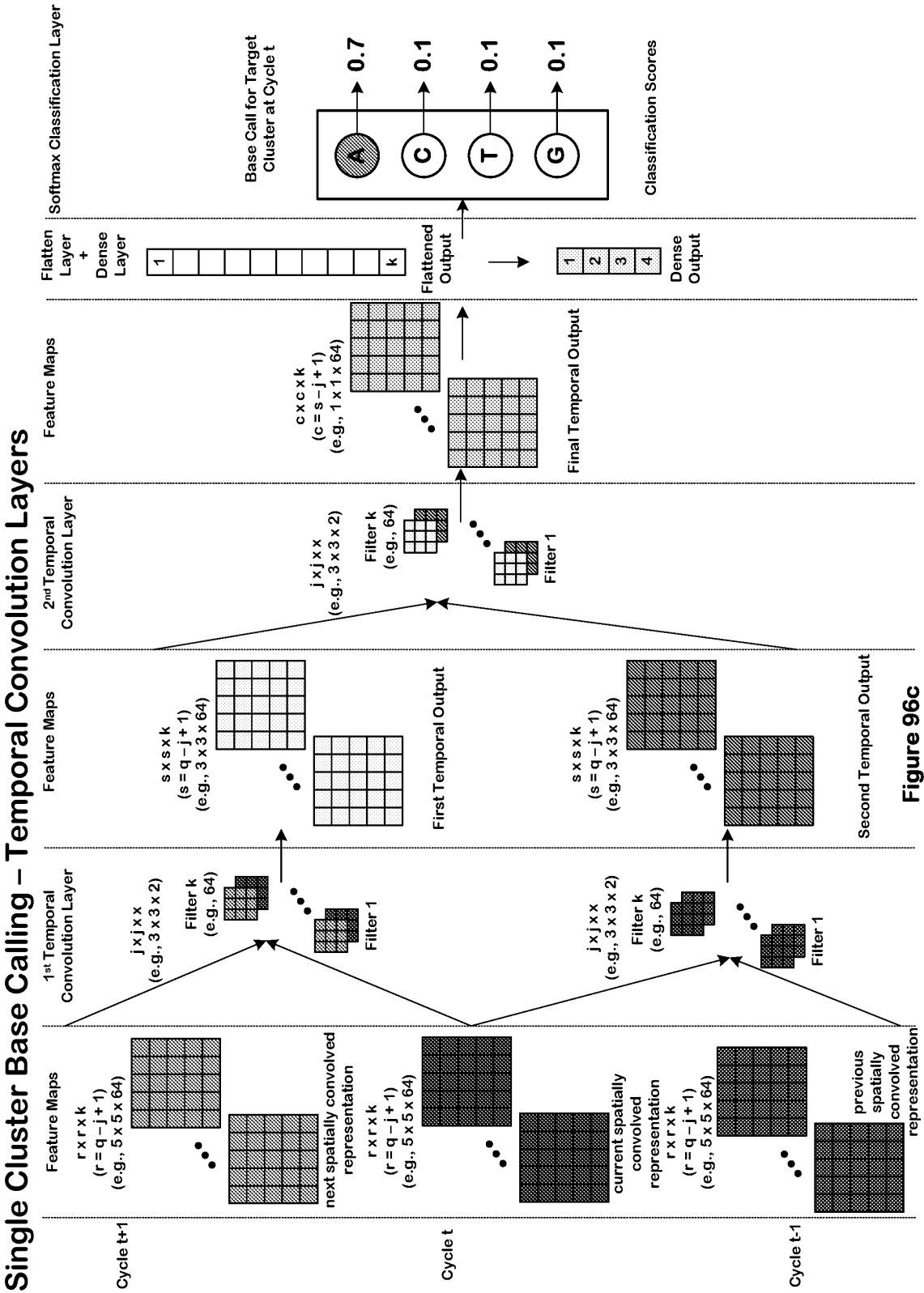


Figure 96c

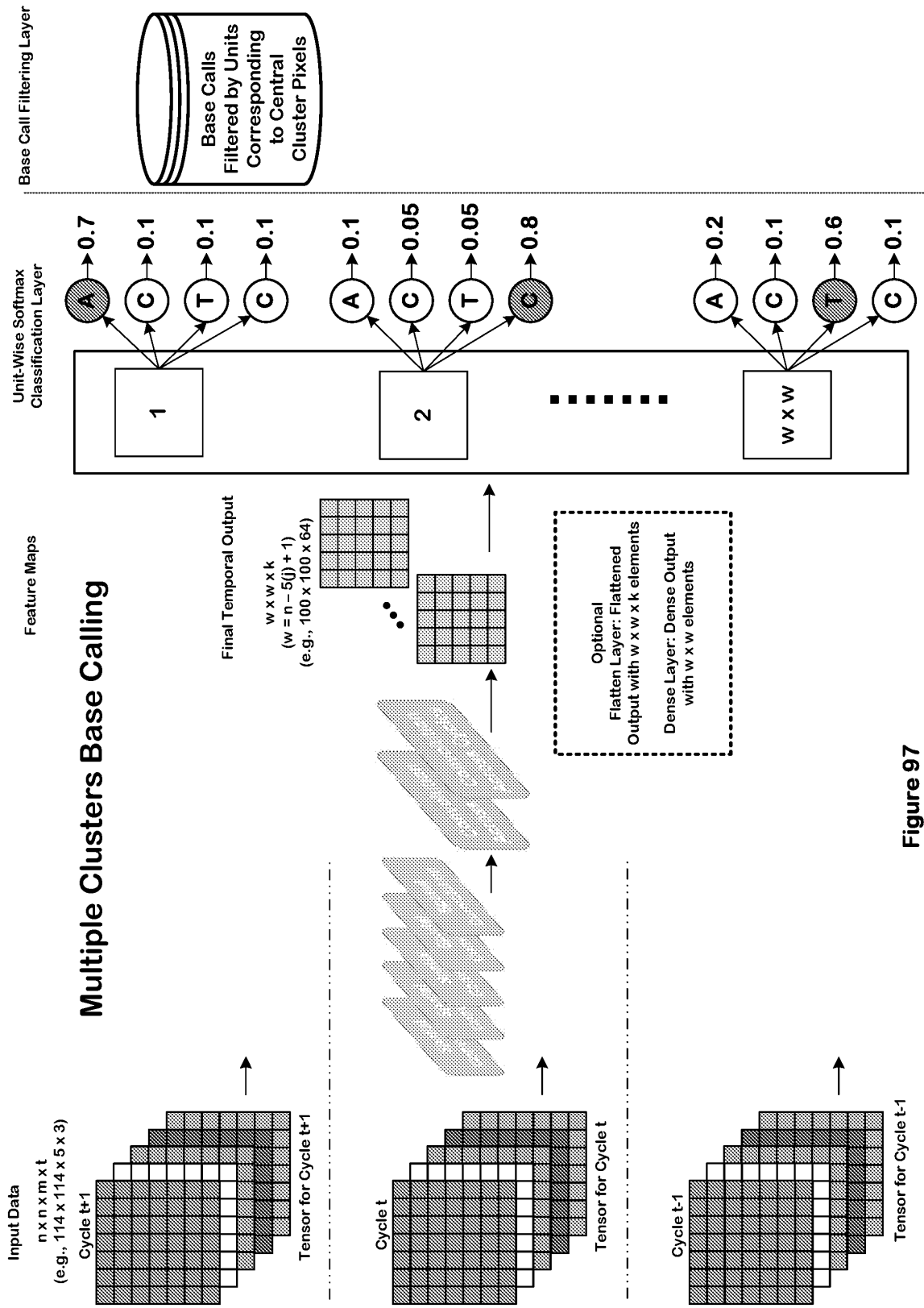


Figure 97

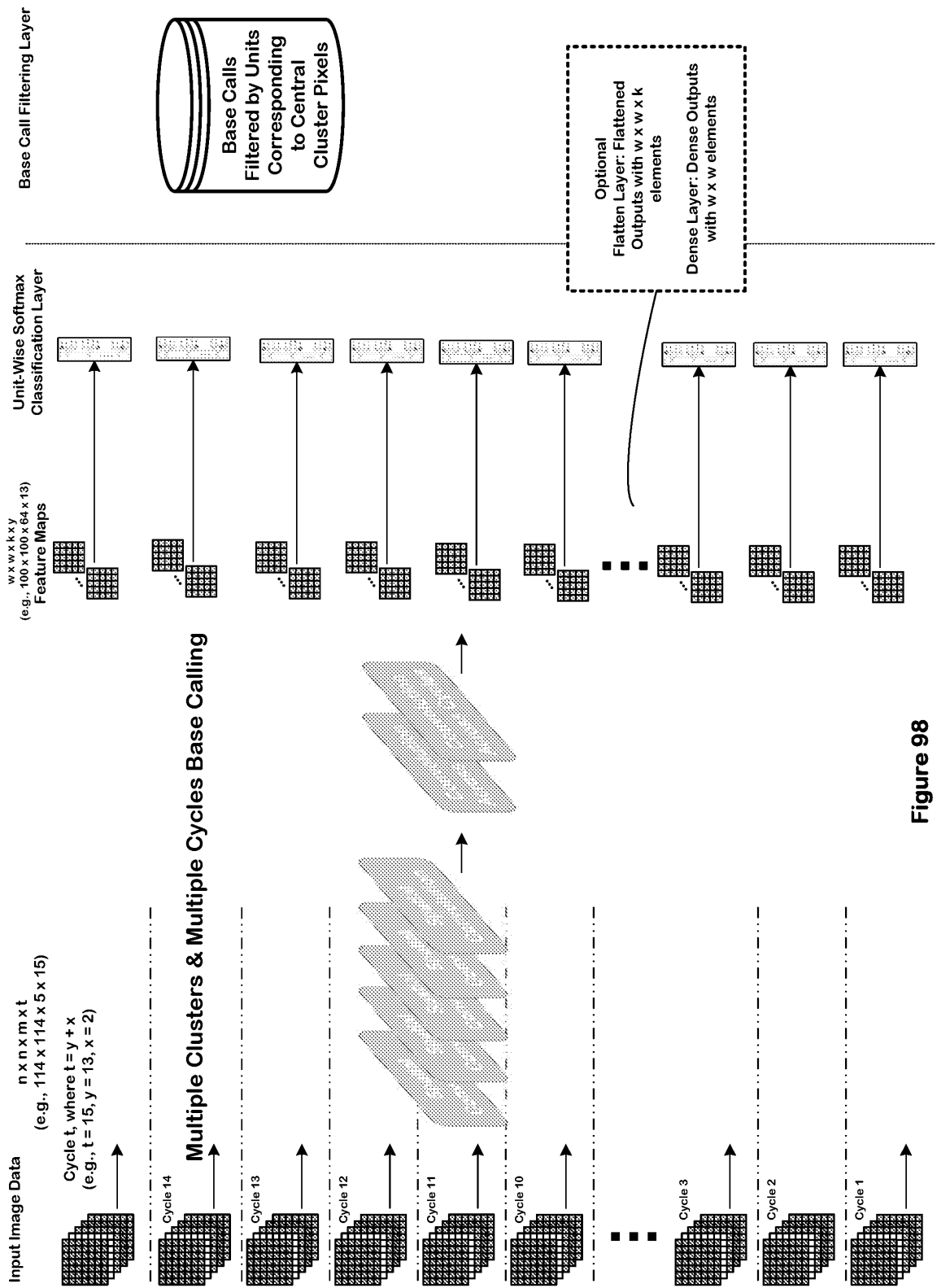


Figure 98

Single Cluster Base Calling: End-to-End Dimensionality Diagram 9900

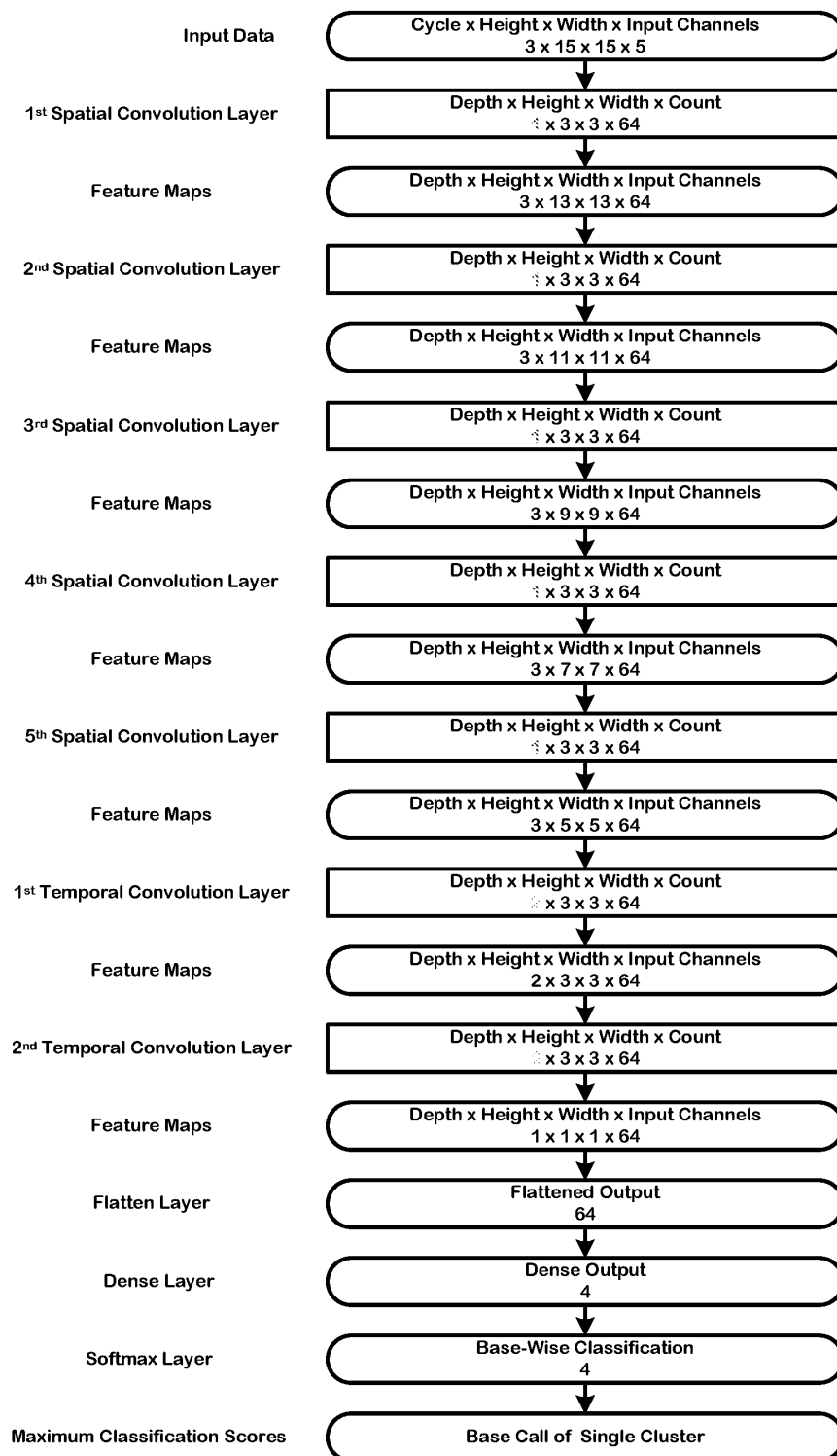


Figure 99

Multiple Clusters Base Calling: End-to-End Dimensionality Diagram 10000

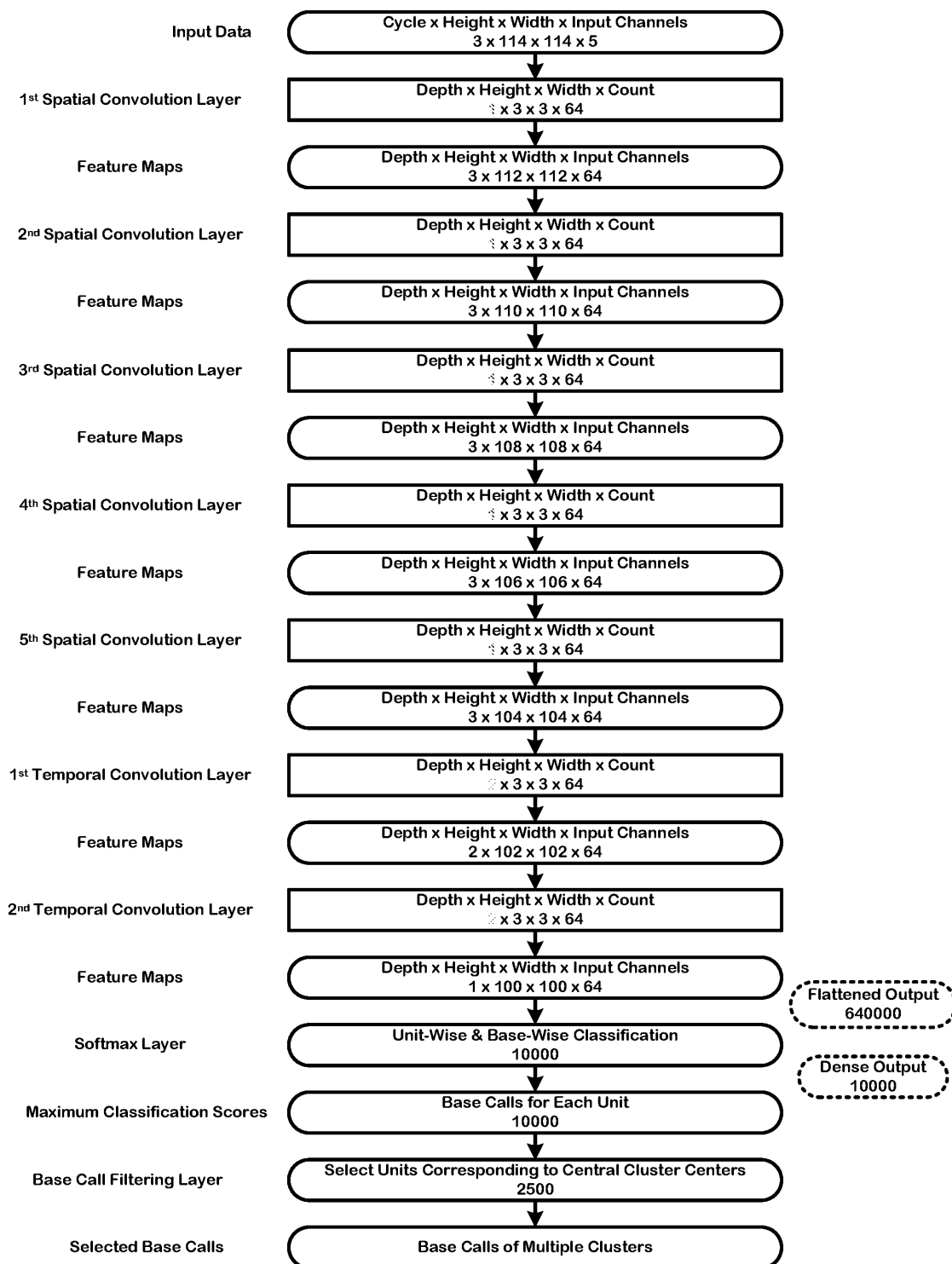
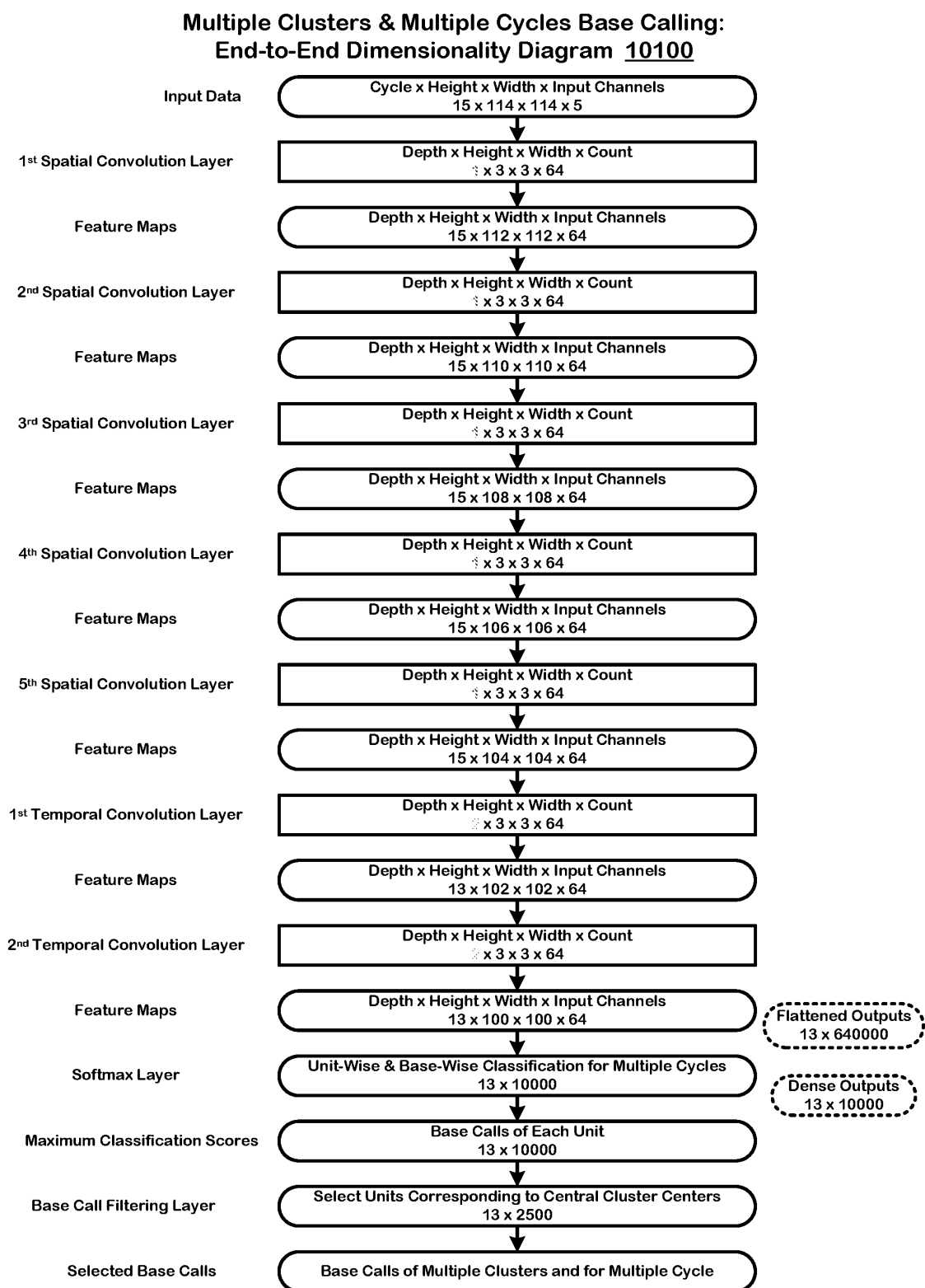
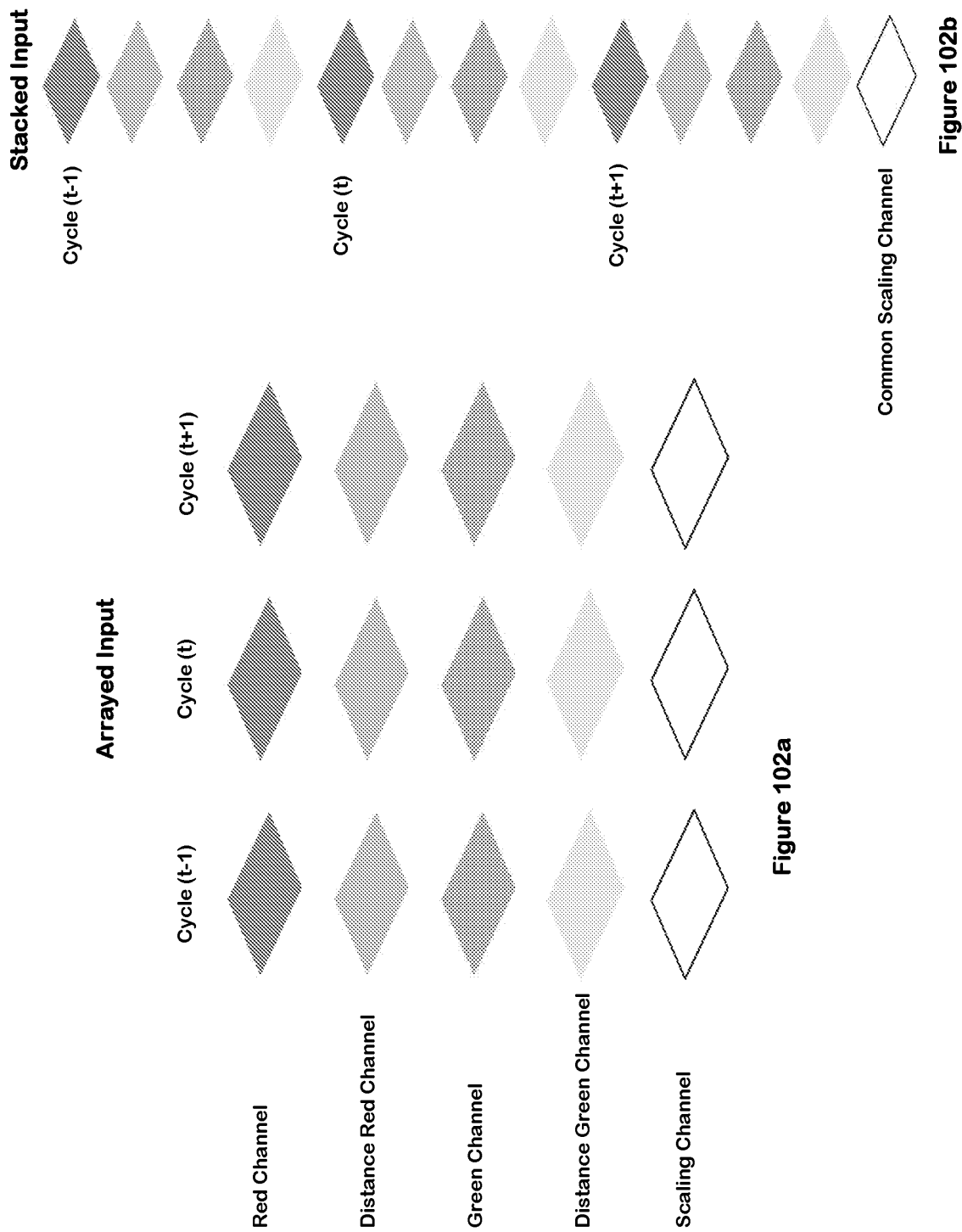
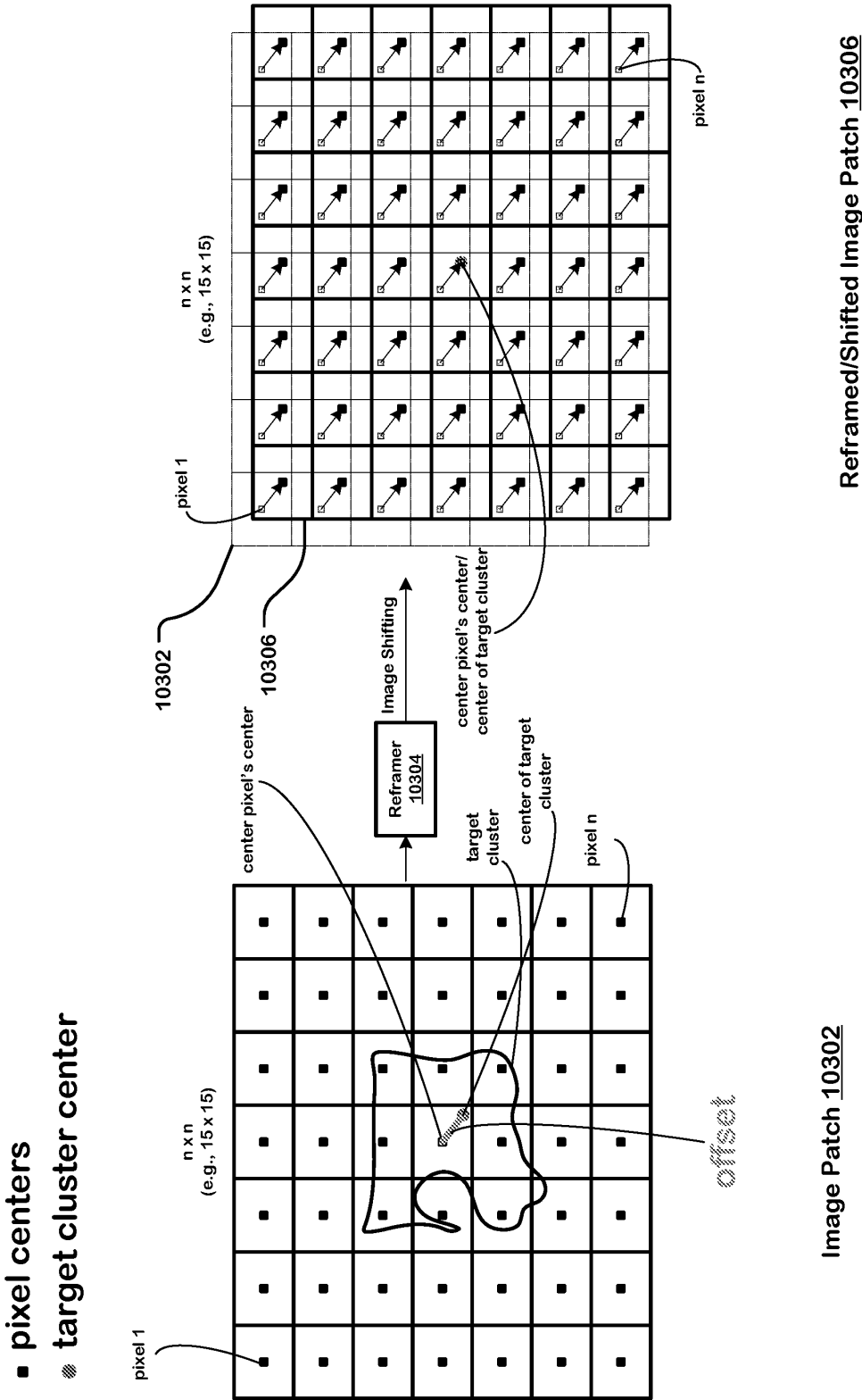


Figure 100

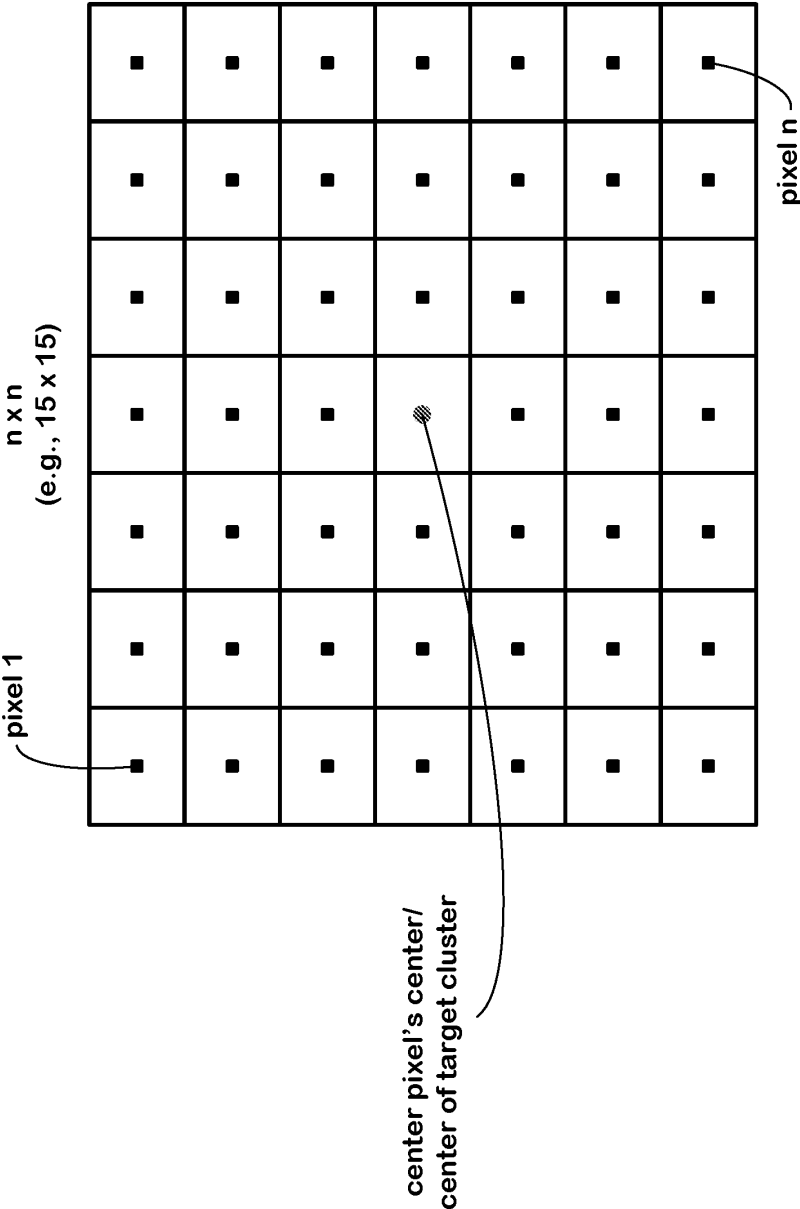
**Figure 101**





Reframing 10300a

Figure 103a



Reframed/Shifted Image Patch 10300b

Figure 103b

Single Cluster Base Calling – Reframing w/o Distance Channels & Specialized CNN 10400

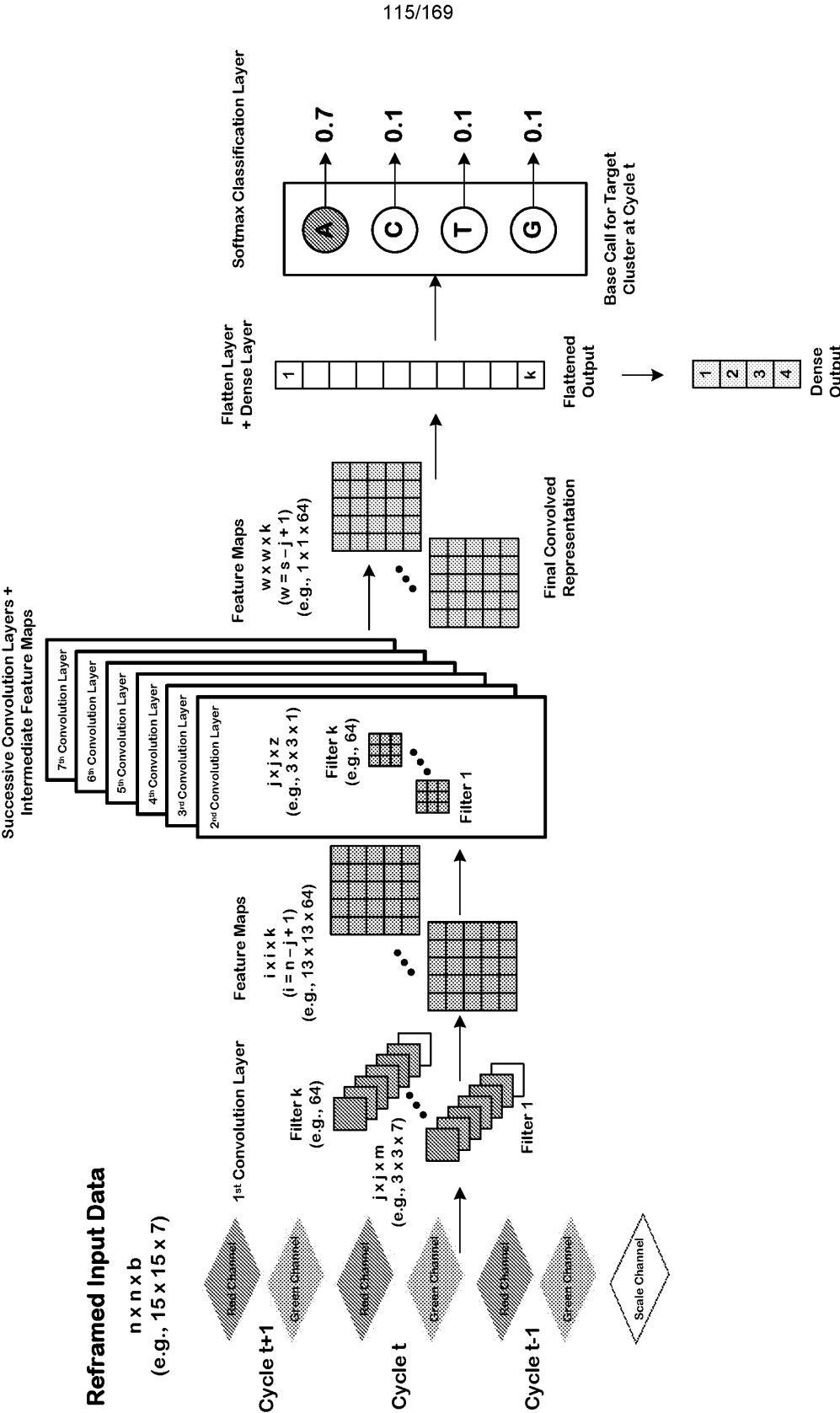


Figure 104

Multiple Clusters Base Calling – w/o the Reframing, Distance Channels & Specialized CNN 10500

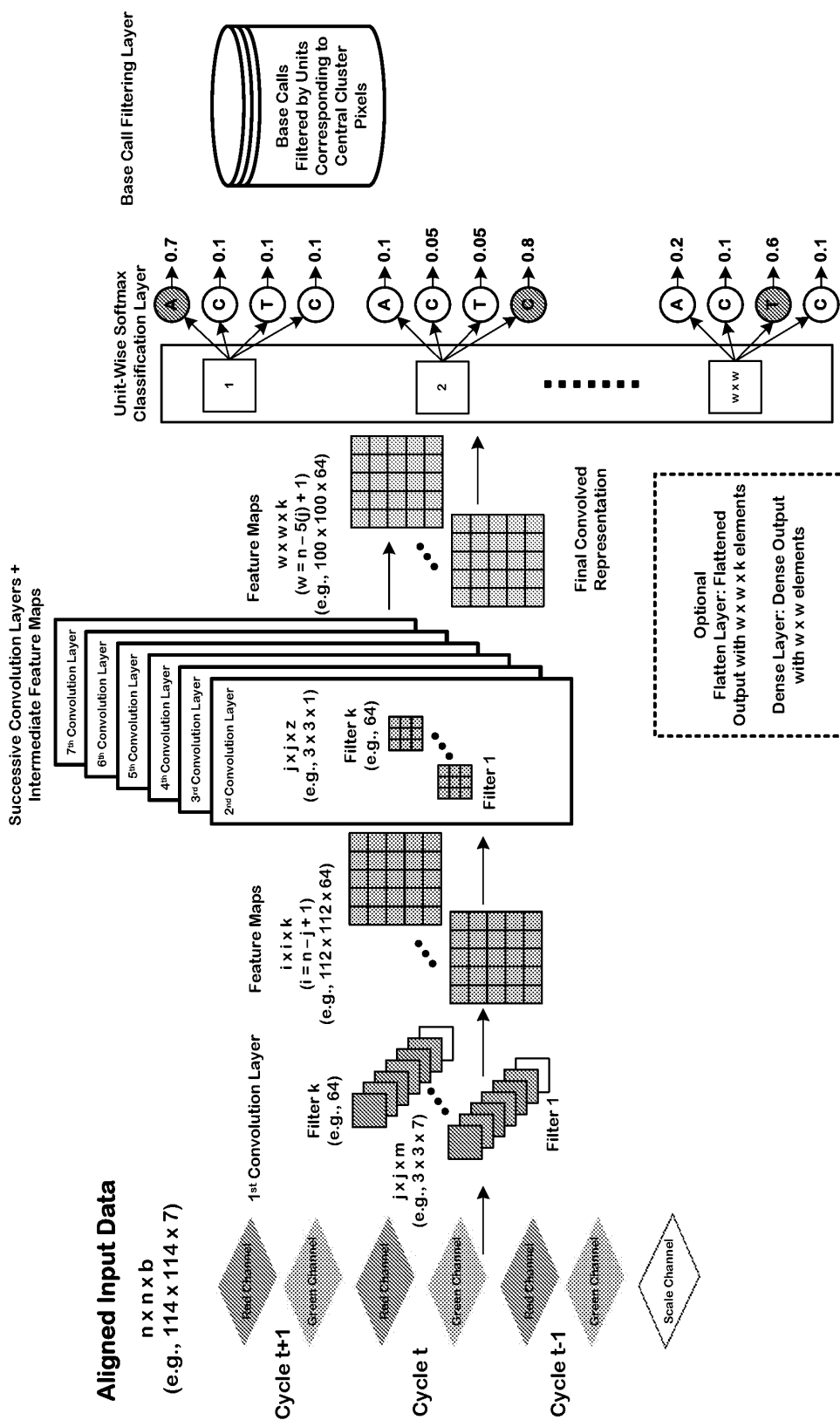


Figure 105

Multiple Clusters & Multiple Cycles Base Calling – w/o the Reframing, Distance Channels & Specialized CNN 10600

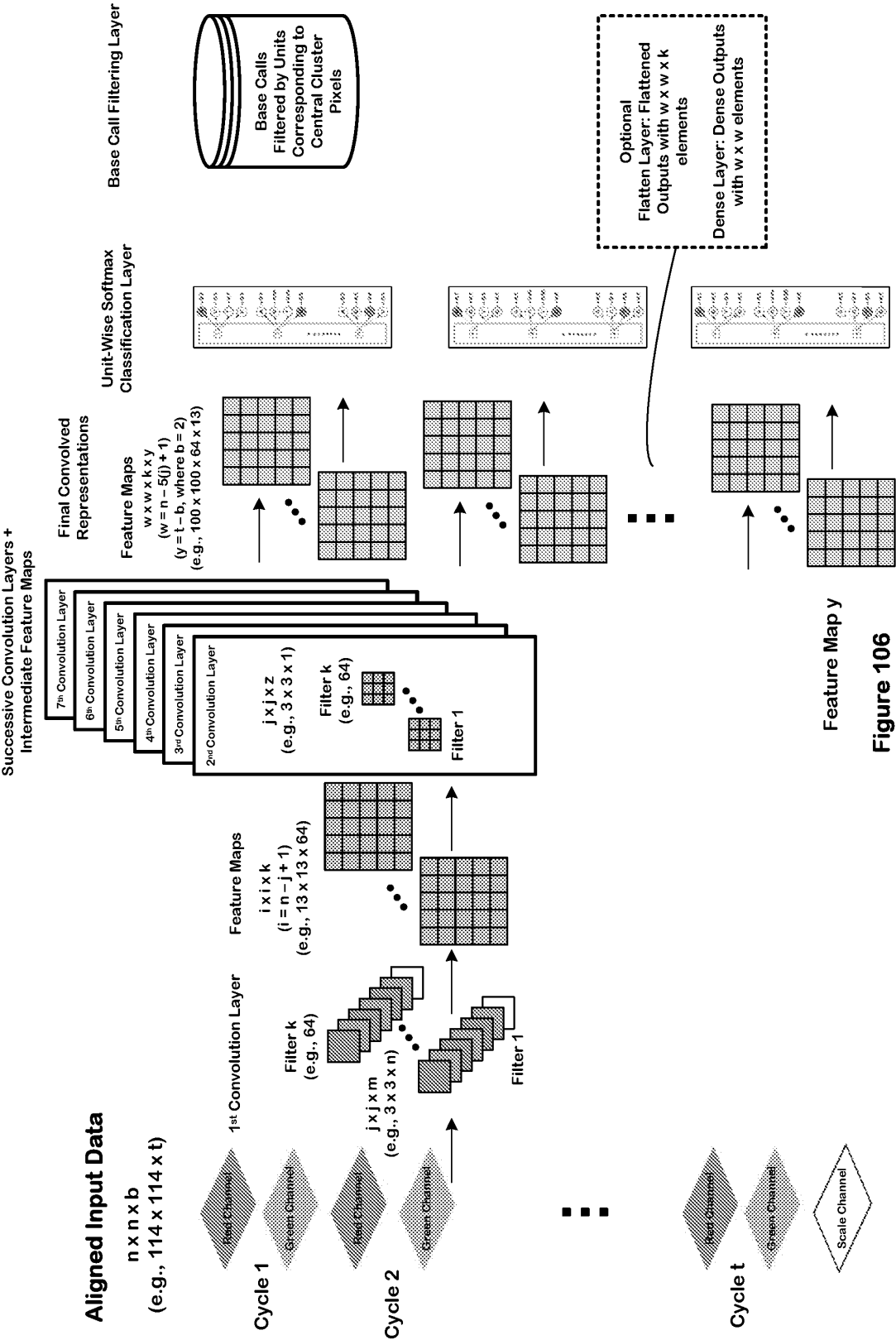


Figure 106

Training the Base Caller 10700

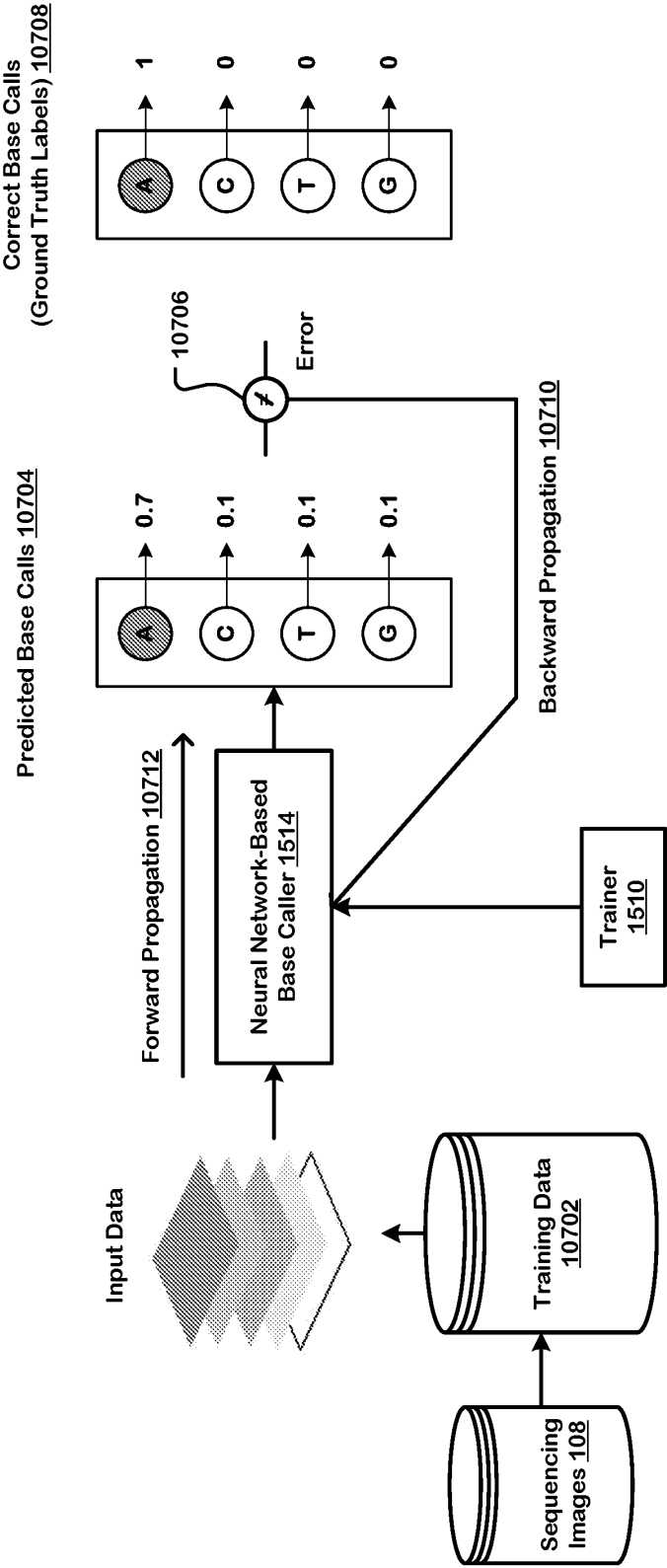


Figure 107

CNN—RNN-based Hybrid Neural Network for Base Calling 10800a

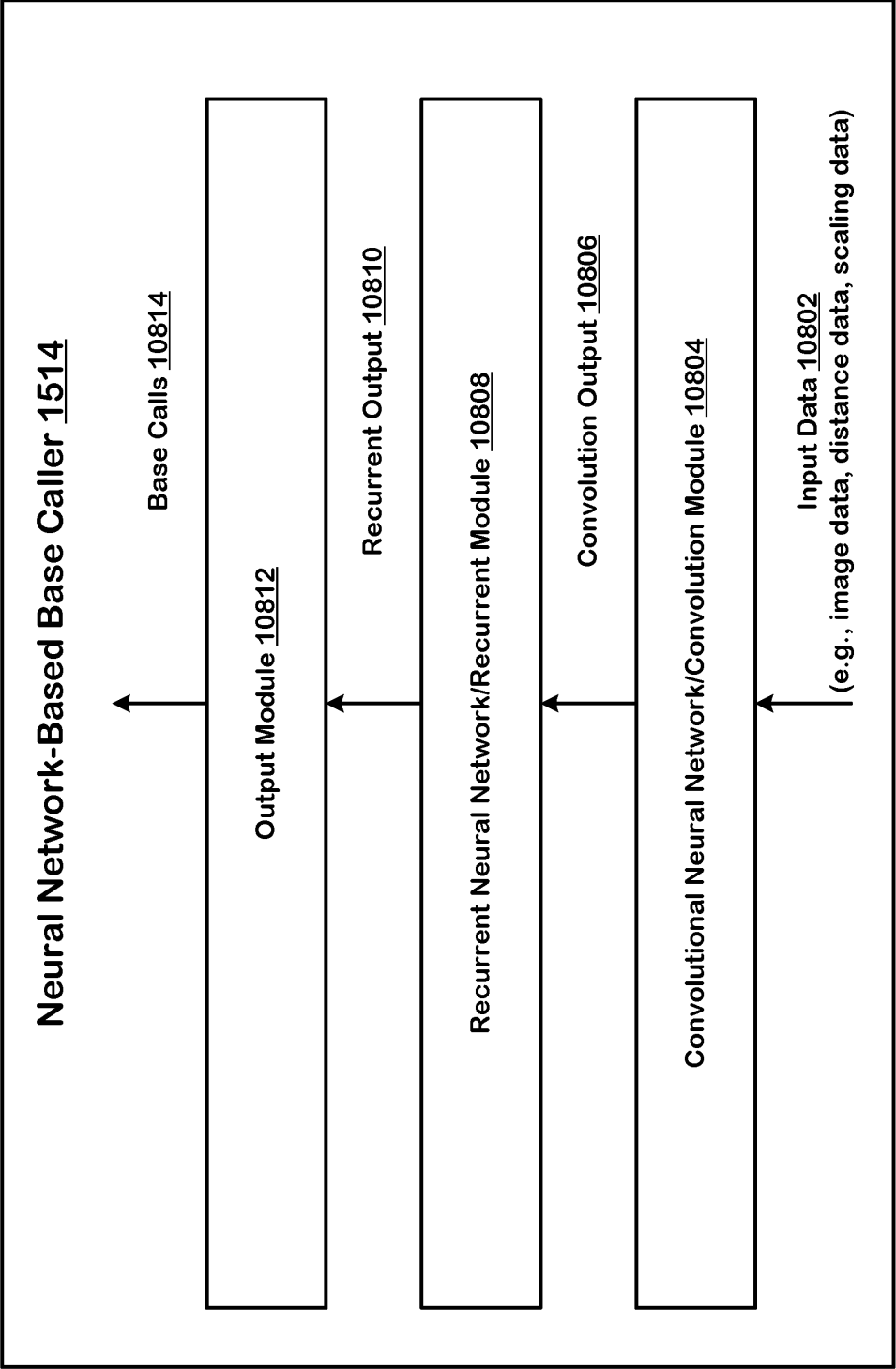


Figure 108a

3D Convolutions 10800b

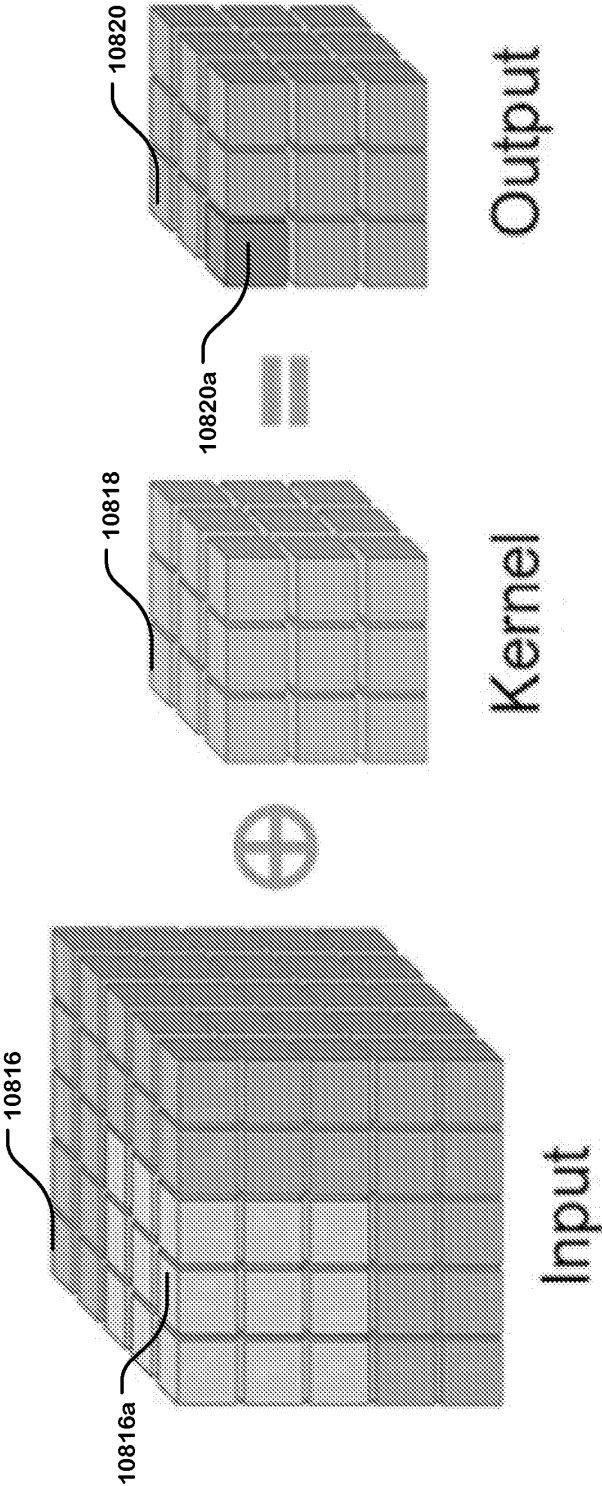


Figure 108b

Cascade of Convolution Layers 10900

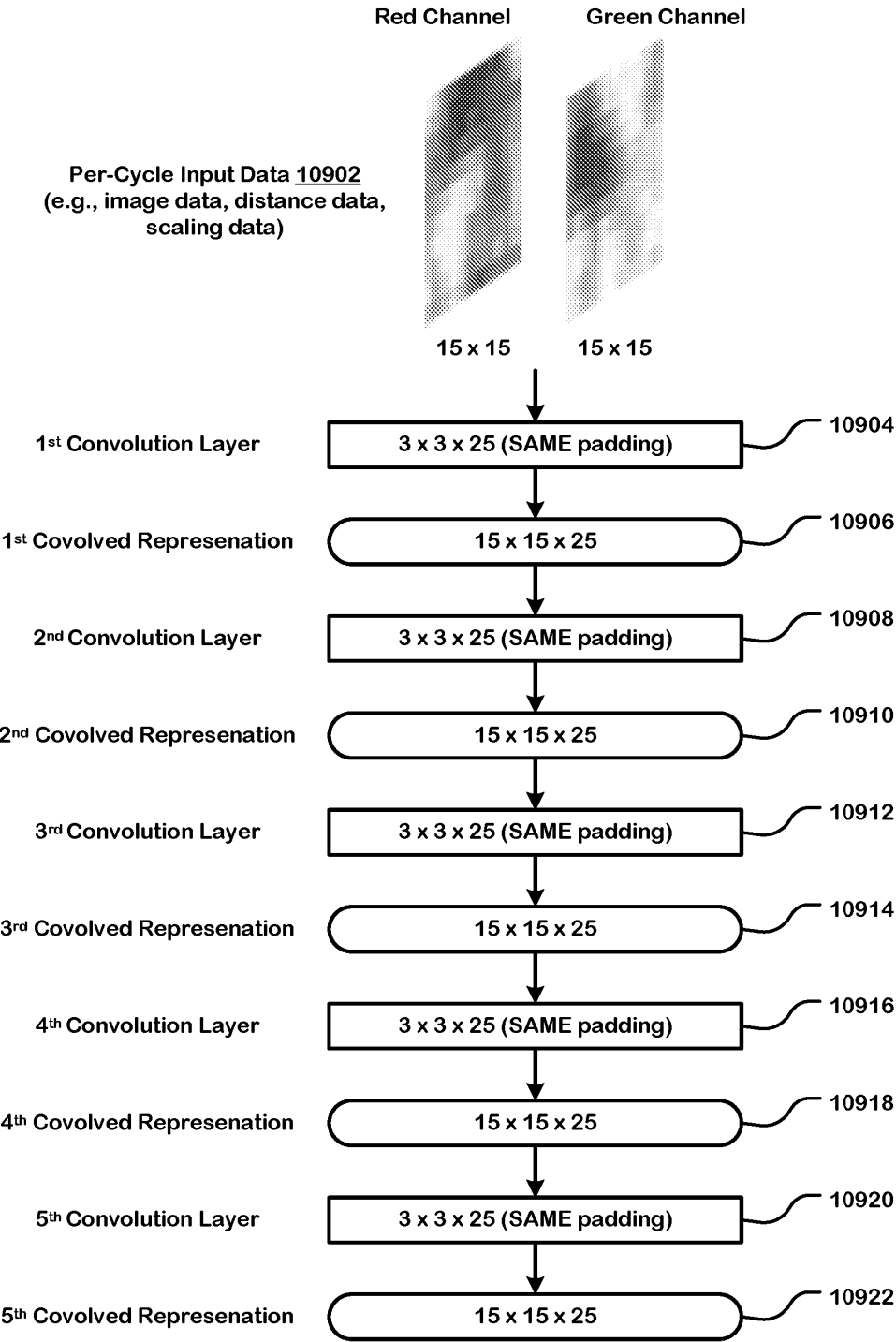


Figure 109

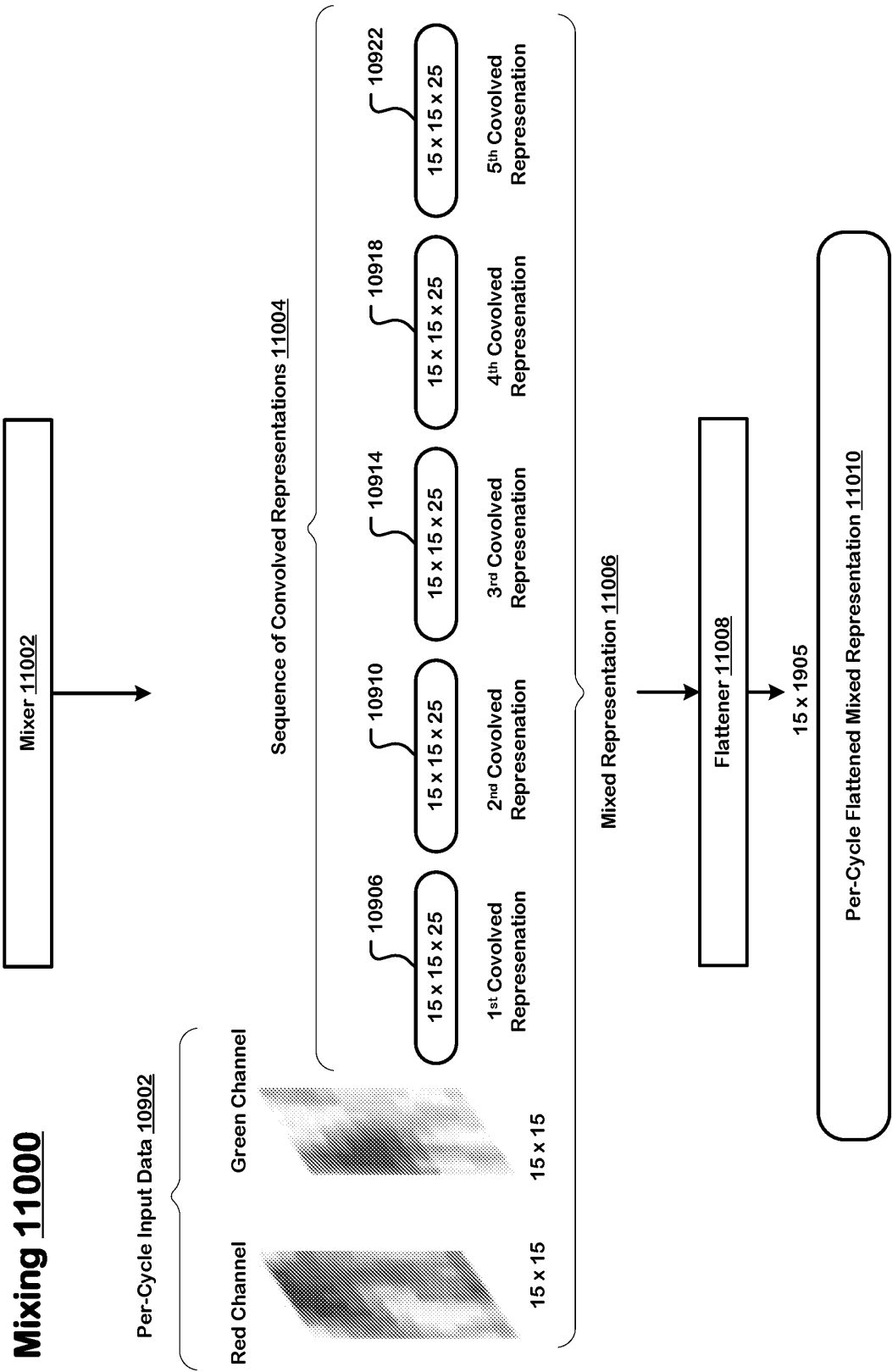


Figure 110

Stack of Flattened Mixed Representations 11100

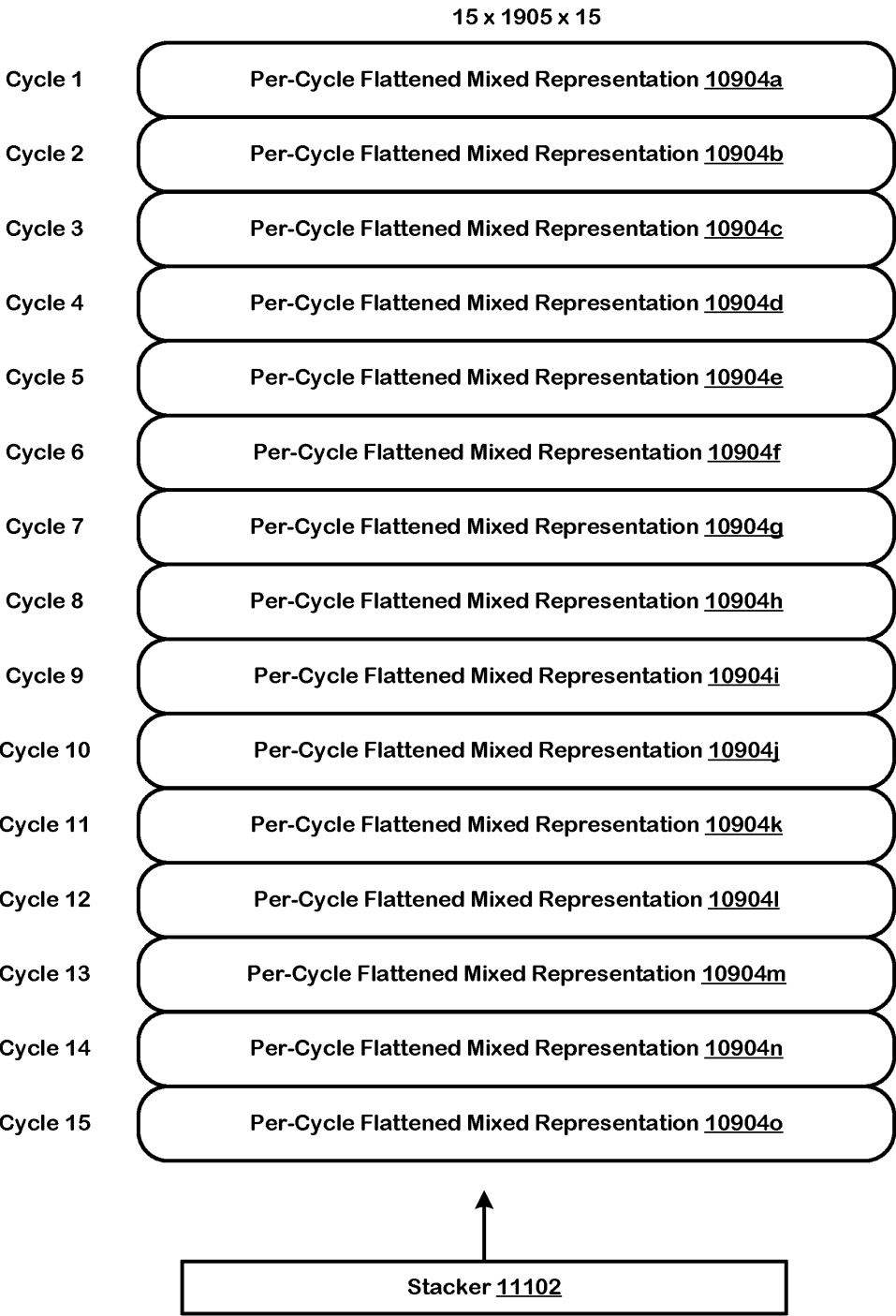
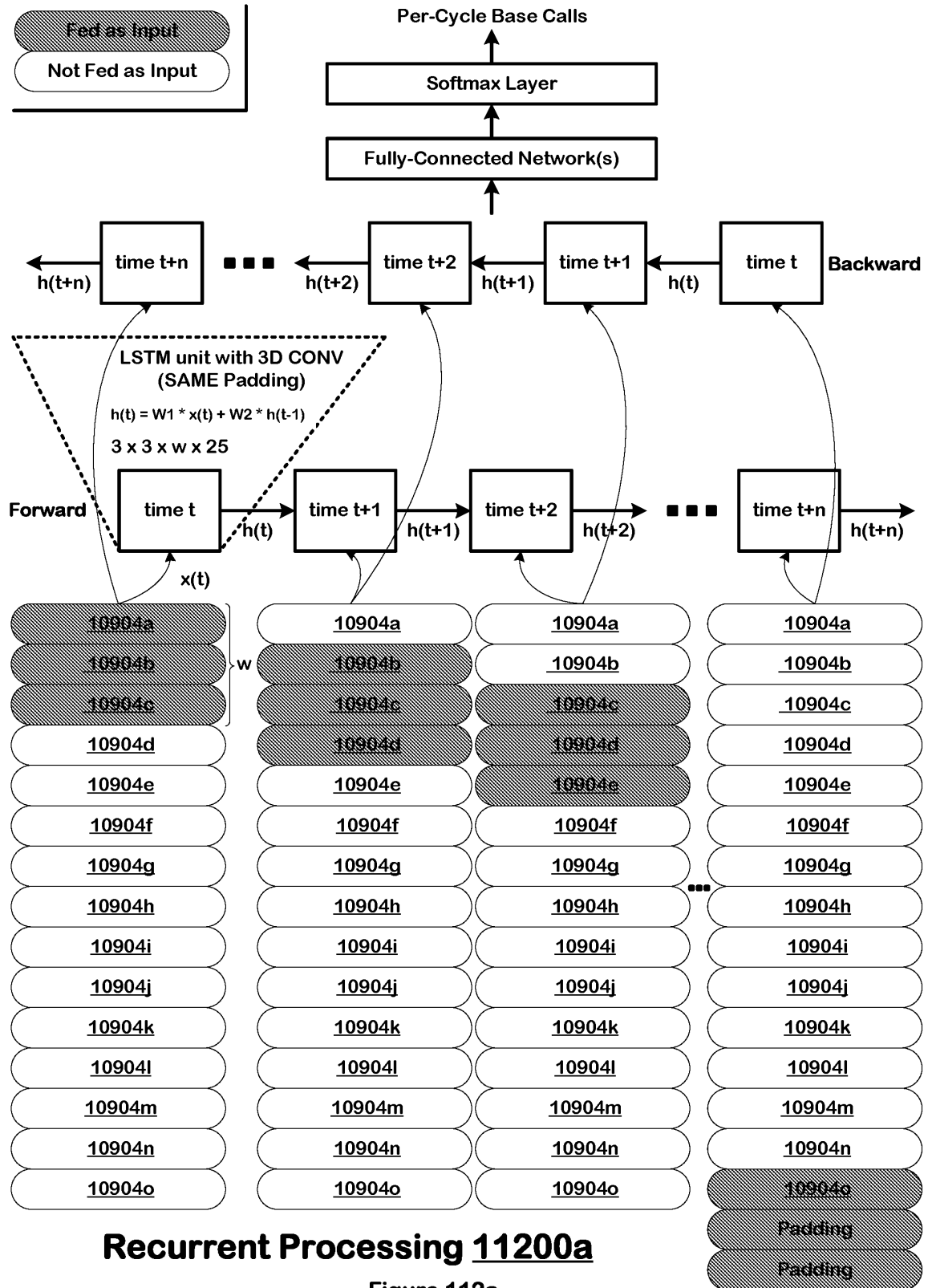
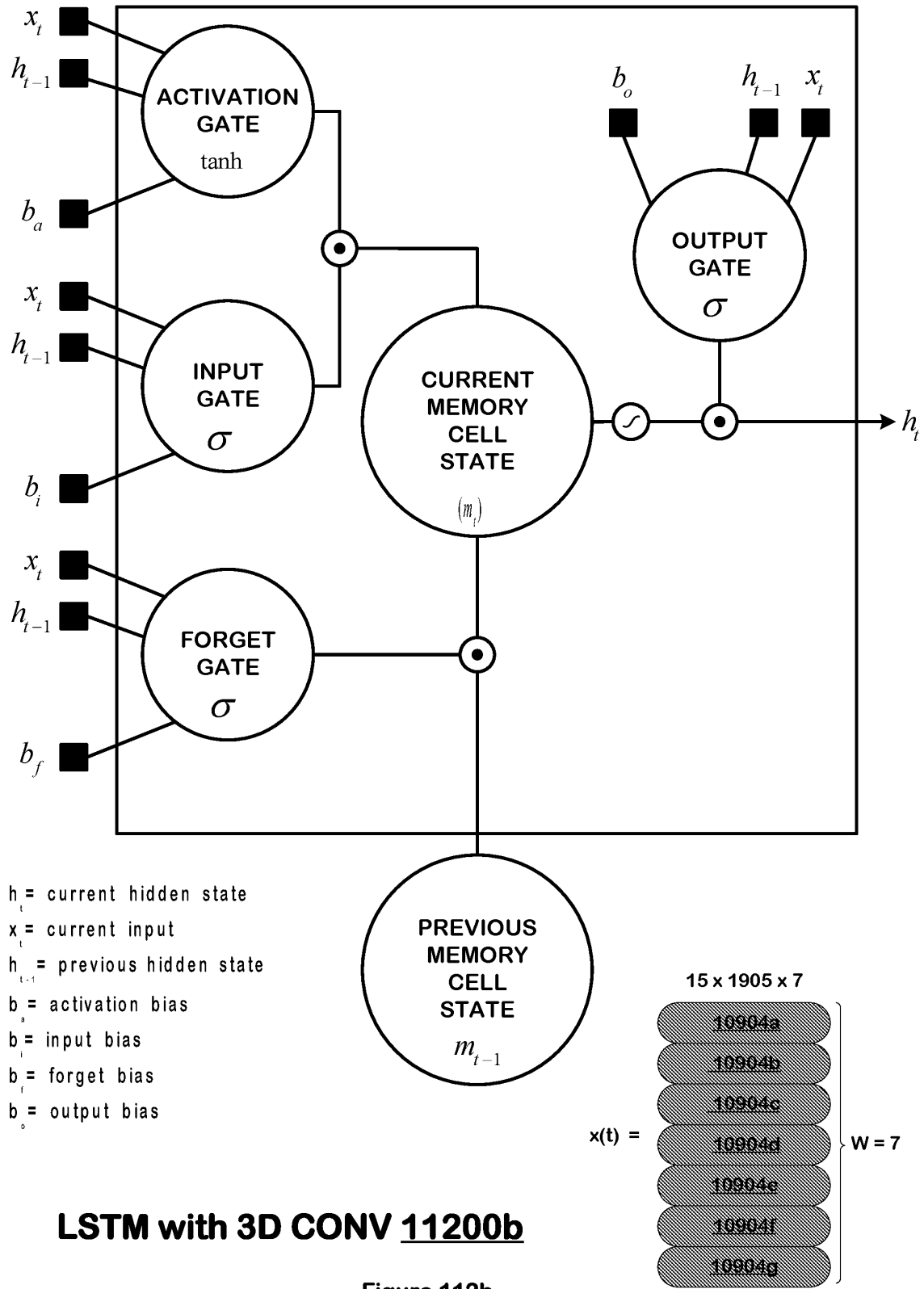


Figure 111

124/169





LSTM with 3D CONV 11200b

Figure 112b

3-mers Balancing in Training Data

- ▶ With more than 300K trainable parameters, network can learn statistics about the genome
- ▶ Balance the number of 3-mers fed to the network to mask genome statistics
- ▶ Use RTA called reads with high mapping score

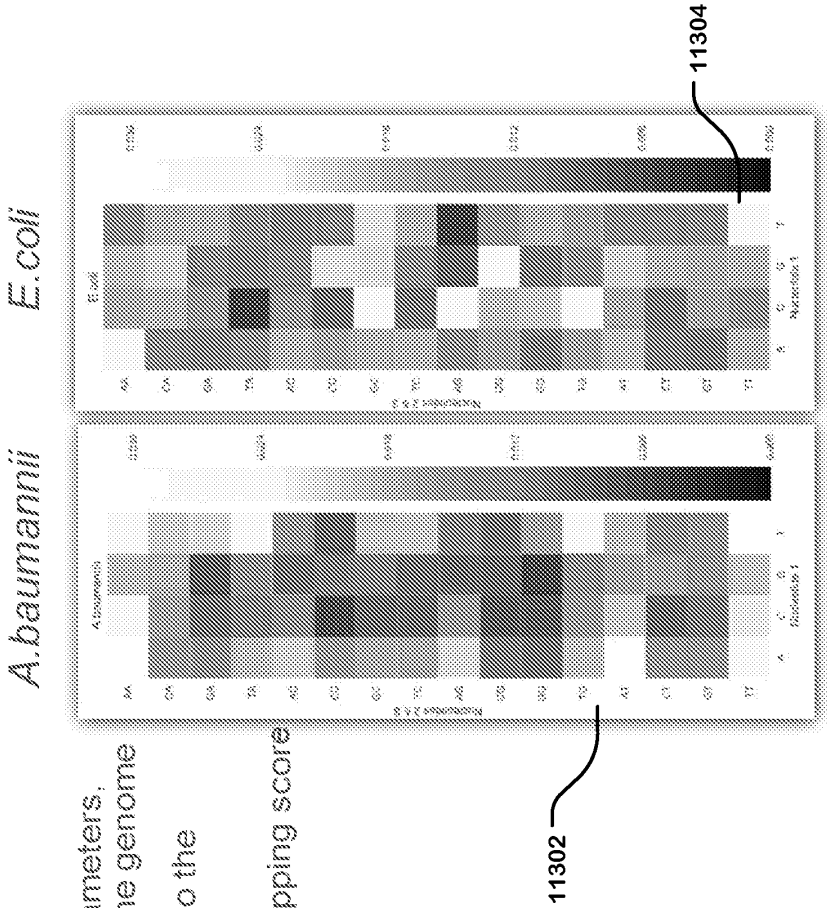
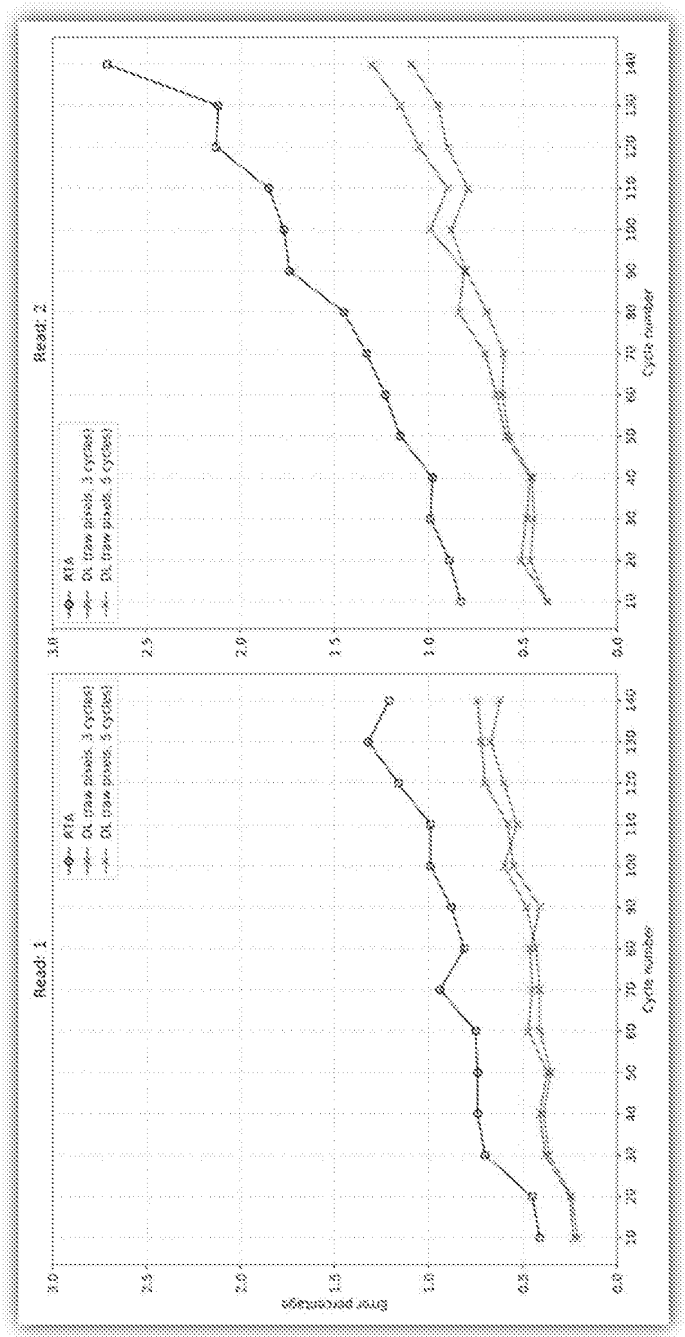


Figure 113

Base Calling using Deep Learning

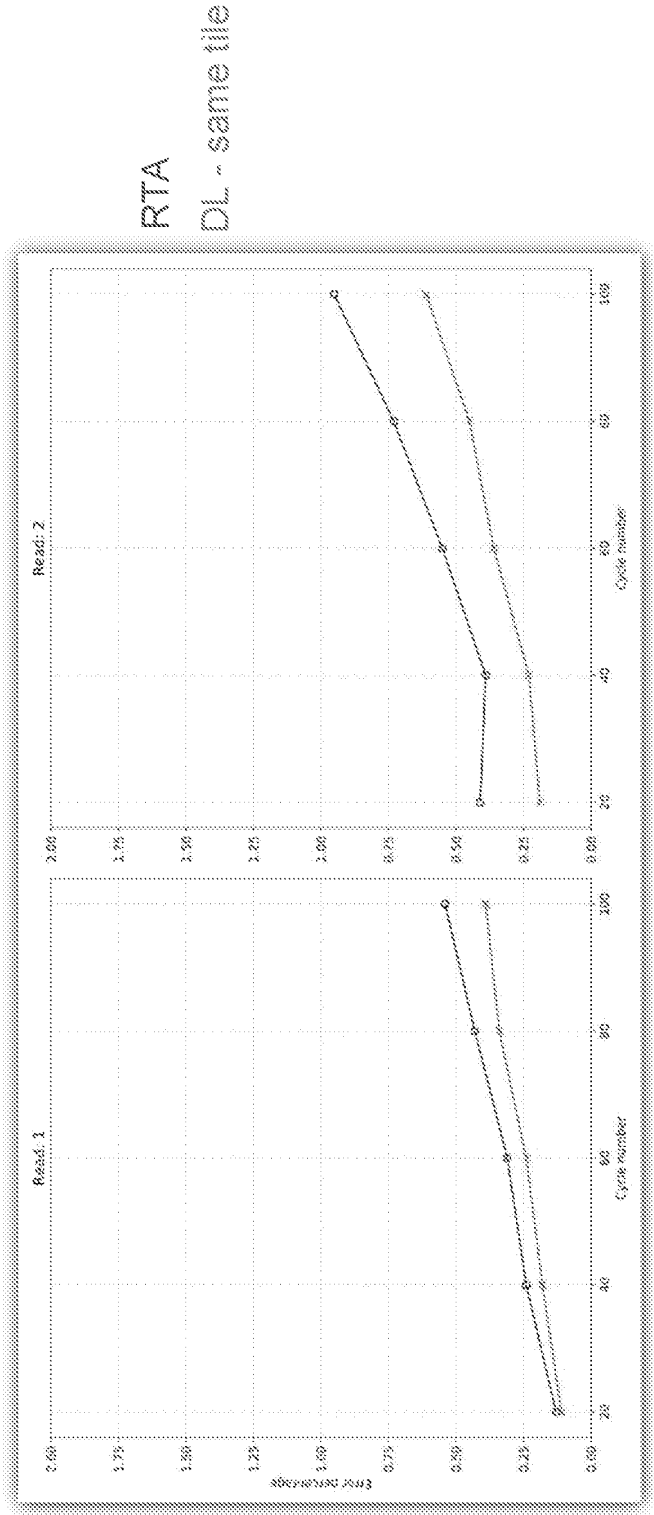


- NextSeq, two channel, random flowcell
- Ensemble of 10 models
- Train and test on the same tile
- ~ 50% reduction in error

Figure 114

Base Calling using Deep Learning

Tile to Tile Generalization

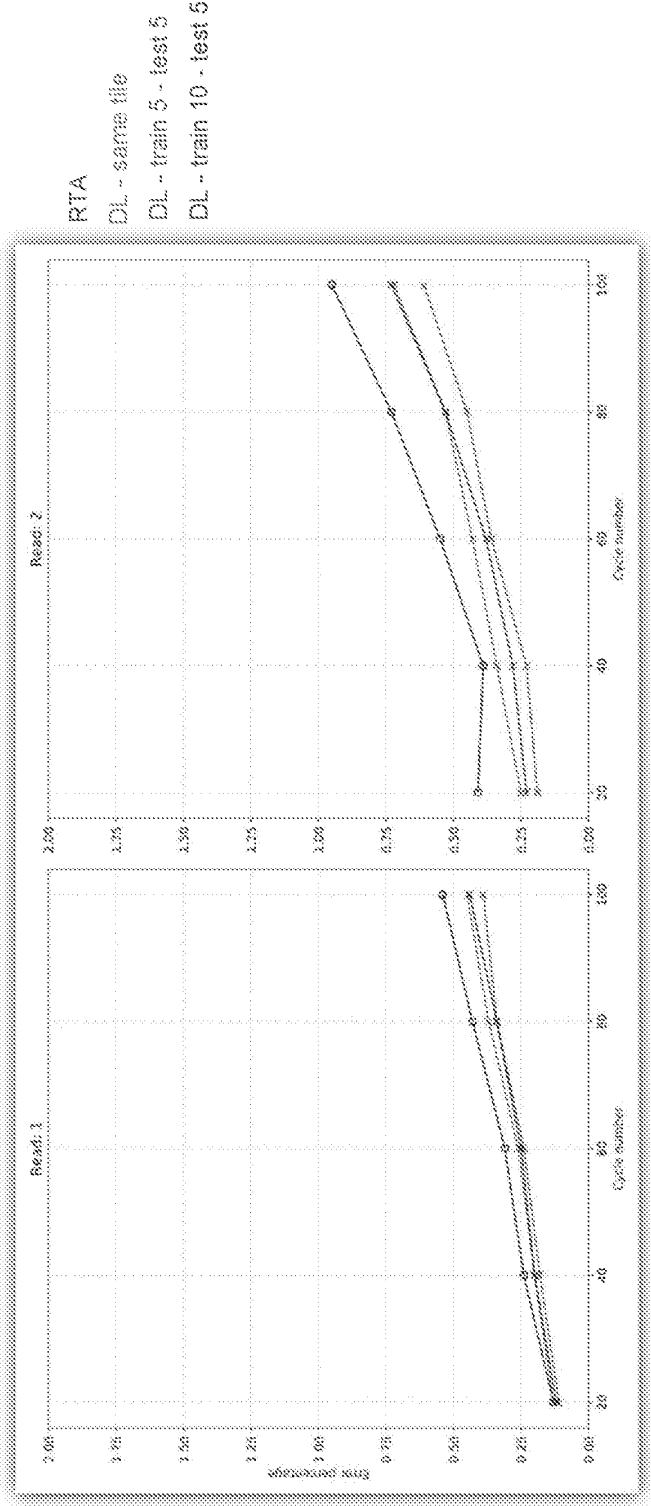


- NextSeq, two channel, random flowcell
- Single Model

Figure 115

Base Calling using Deep Learning

Tile to Tile Generalization

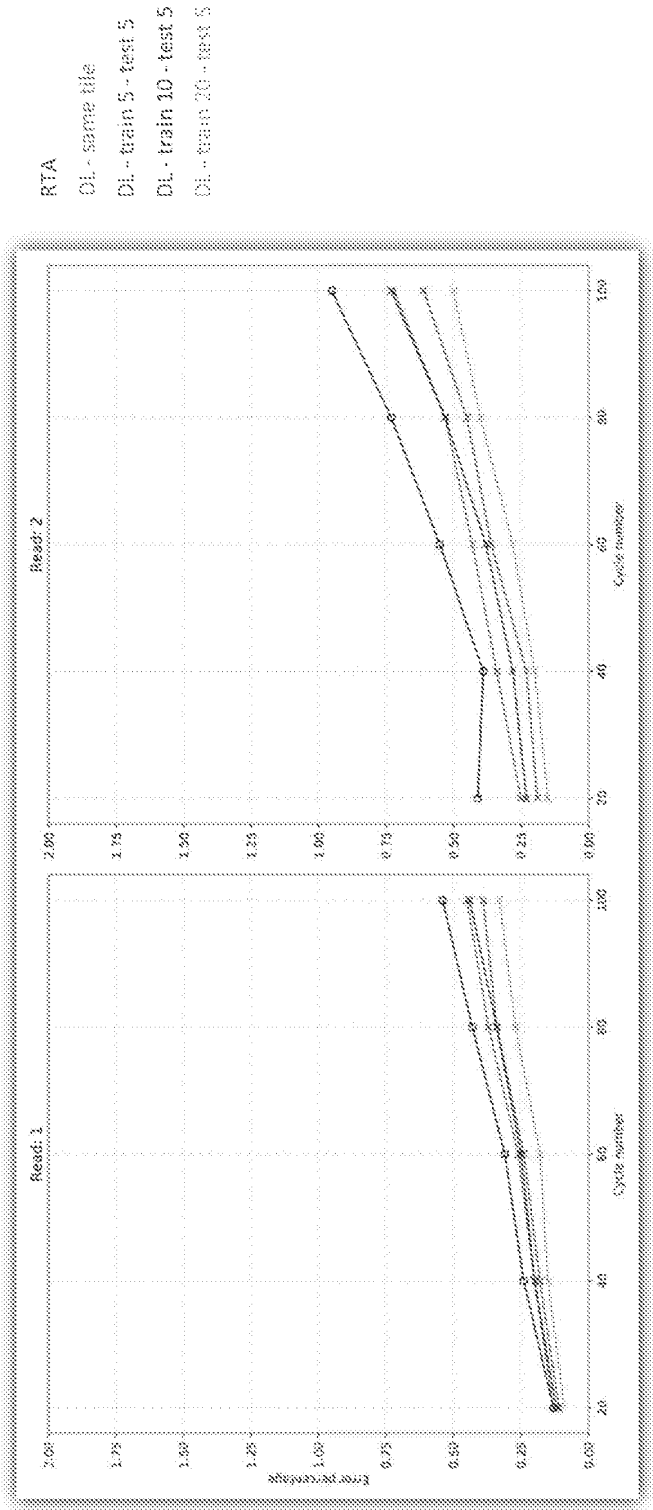


- NextSeq, two channel, random flowcell
- Single Model

Figure 116

Base Calling using Deep Learning

Tile to Tile Generalization



- NextSeq, two channel, random flowcell
- No ensemble - single model

Figure 117

PATCH SIZE EFFECT

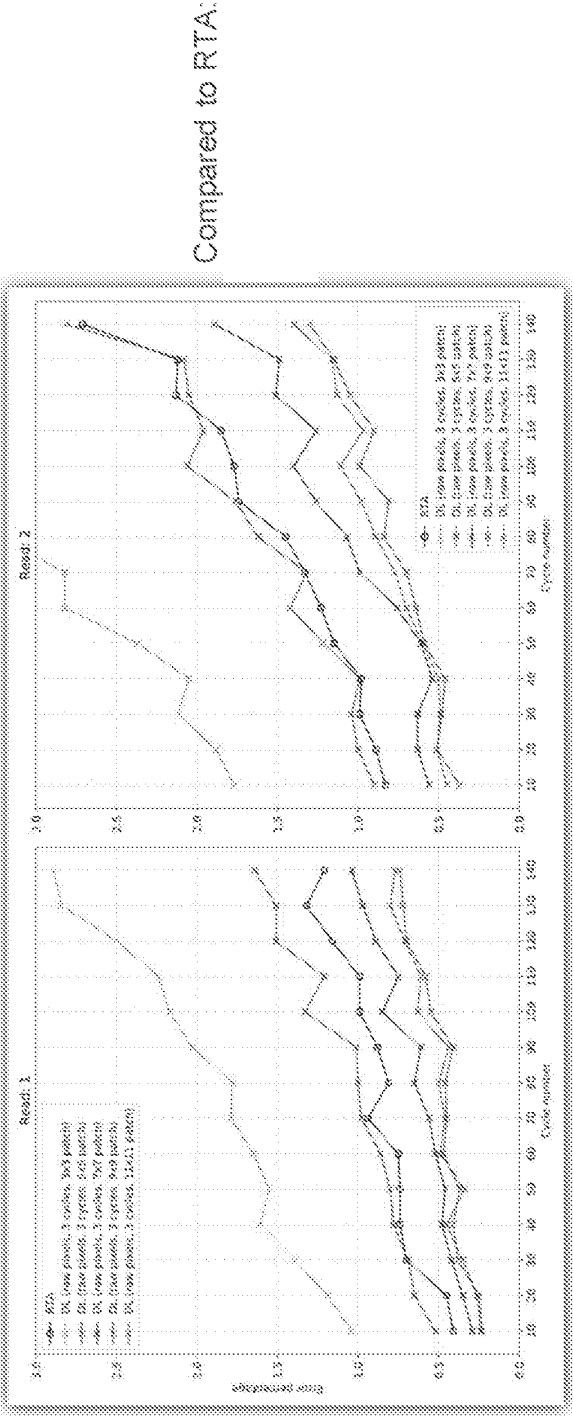
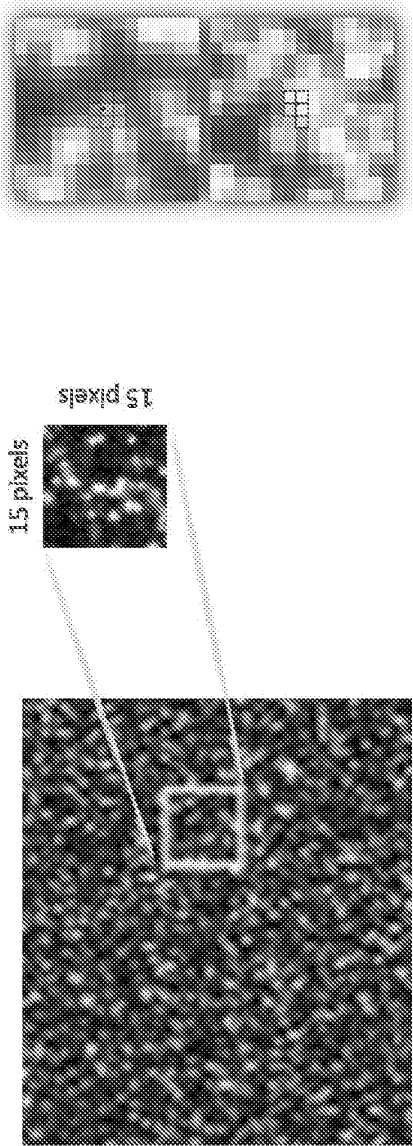


Figure 118

132/169

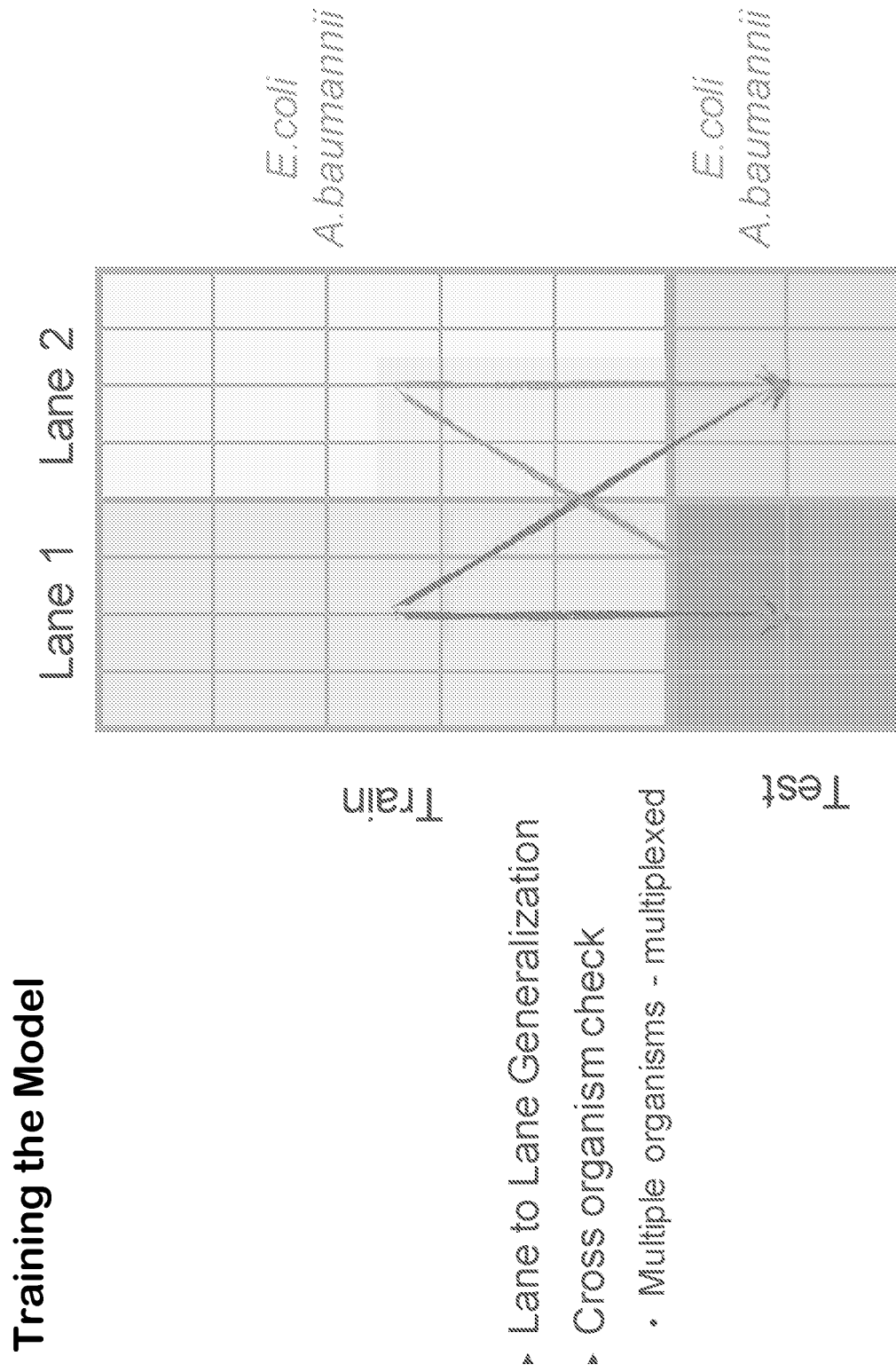


Figure 119

Test: *A. baumannii*

Test: *E. coli*

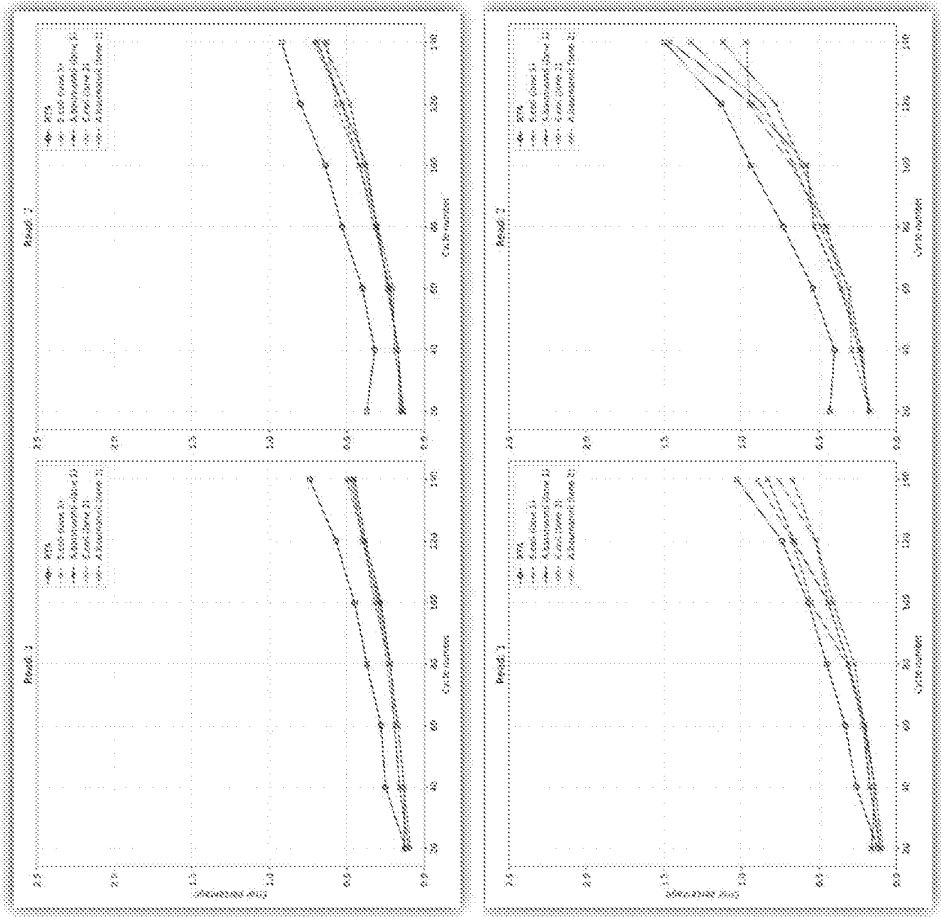


Figure 120

Test: *A. baumannii*
(lane 1)

Test: *E. coli*
(lane 1)

Between 50-80%
error reduction

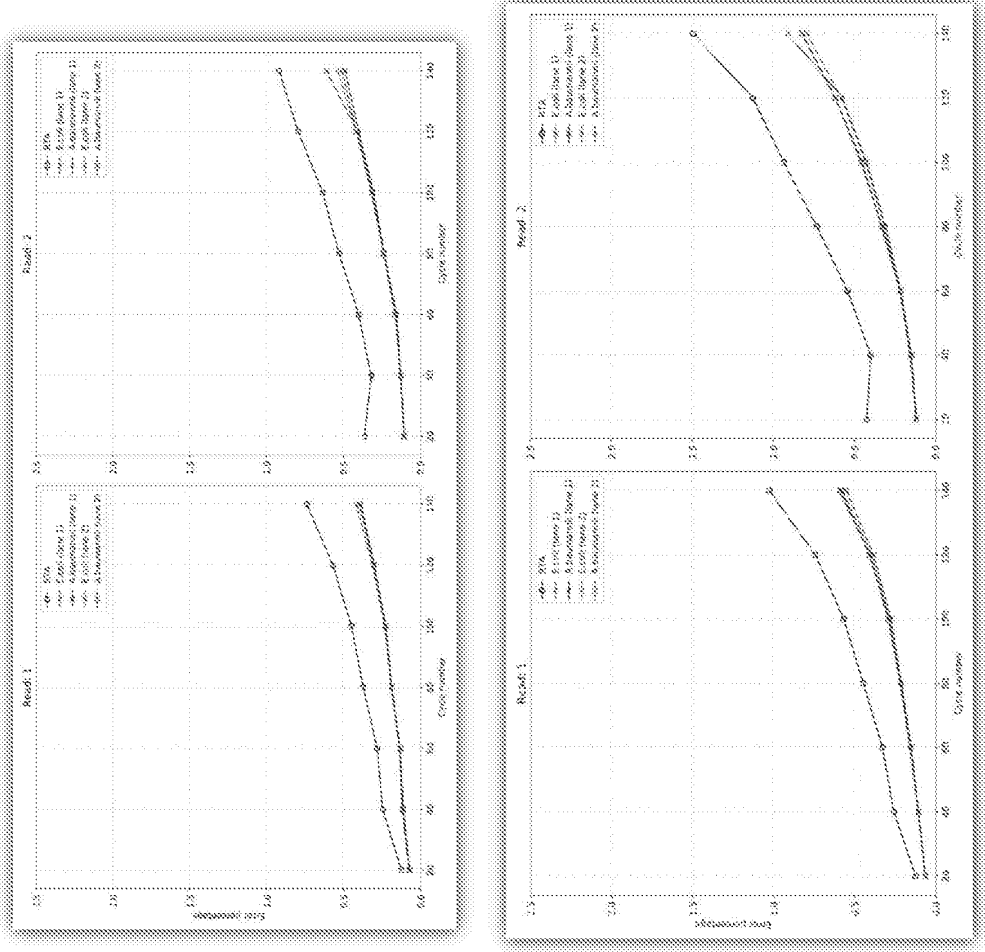
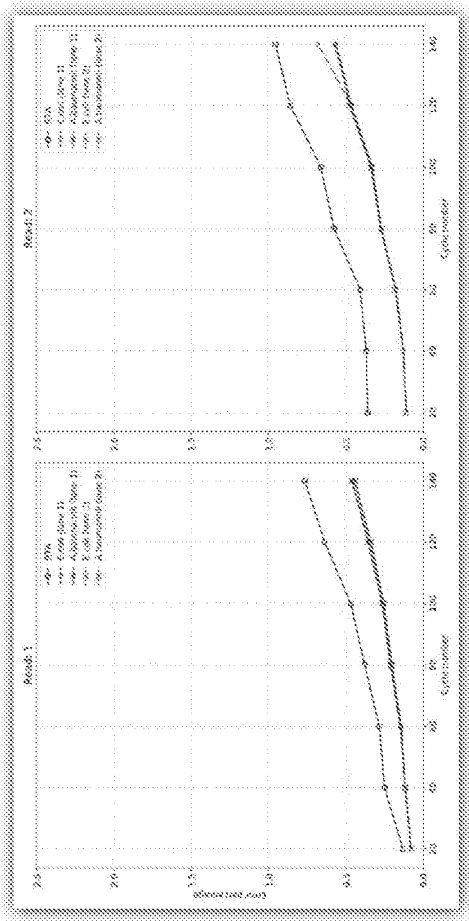


Figure 121

Test: *A. baumannii*
(lane 2)



Test: *E. coli*
(lane 2)

Between 50-80%
error reduction

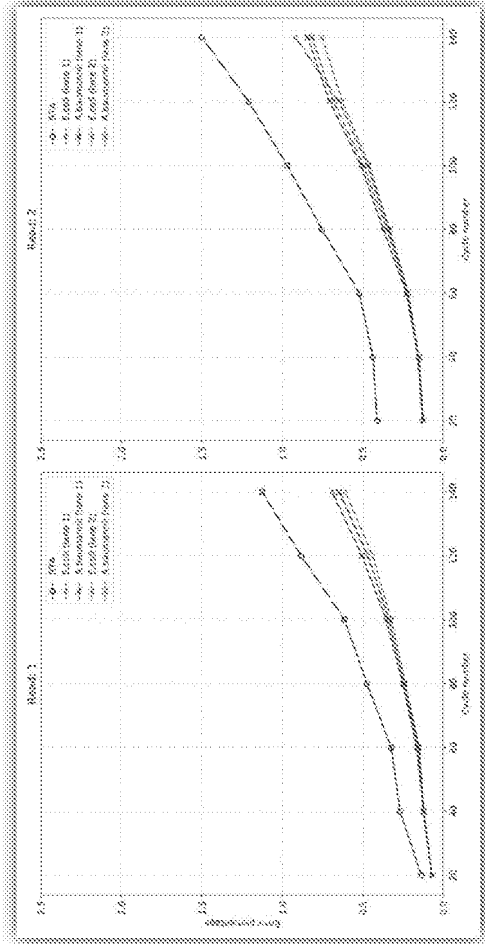


Figure 122

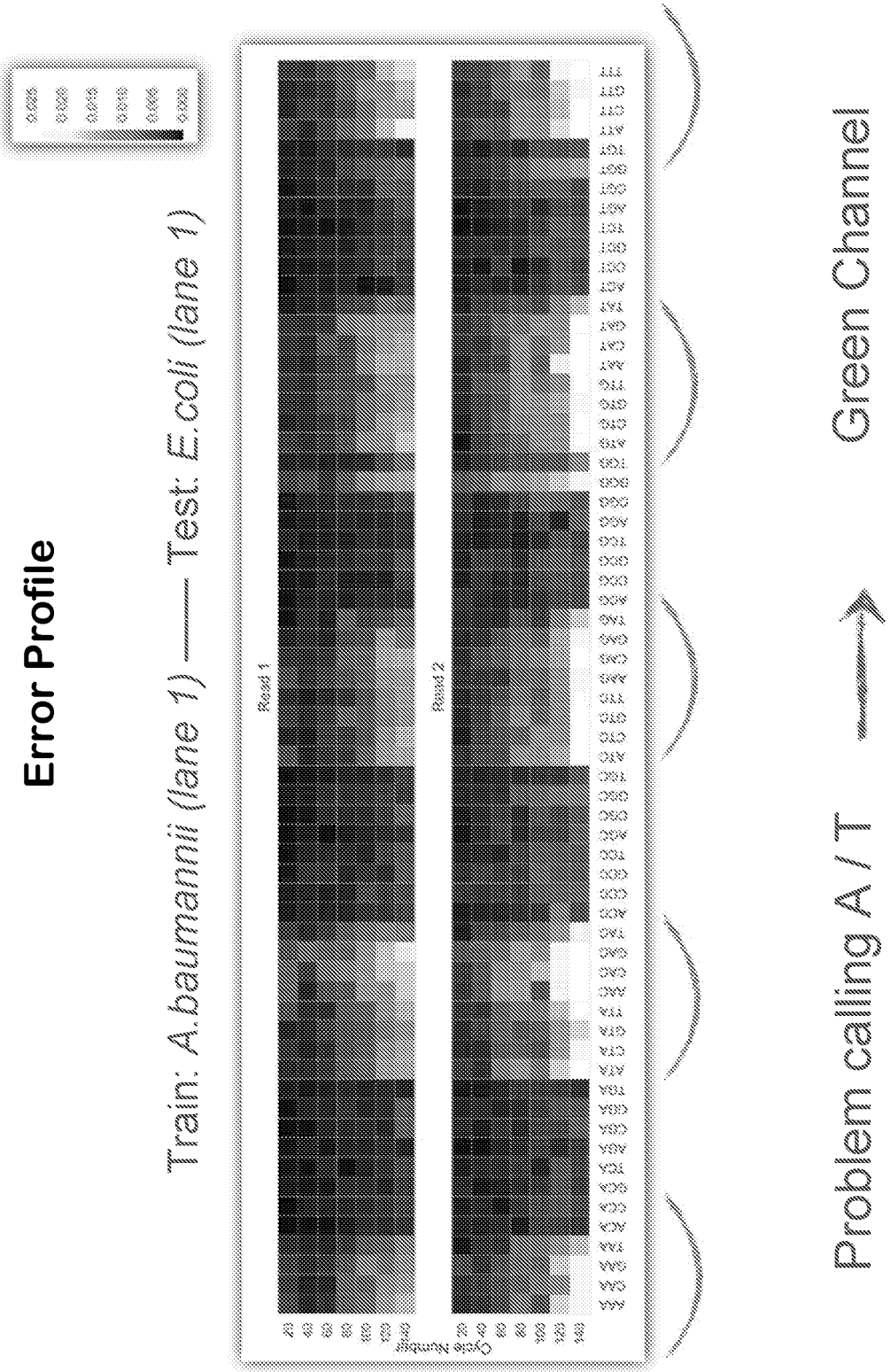


Figure 123

Source of Error

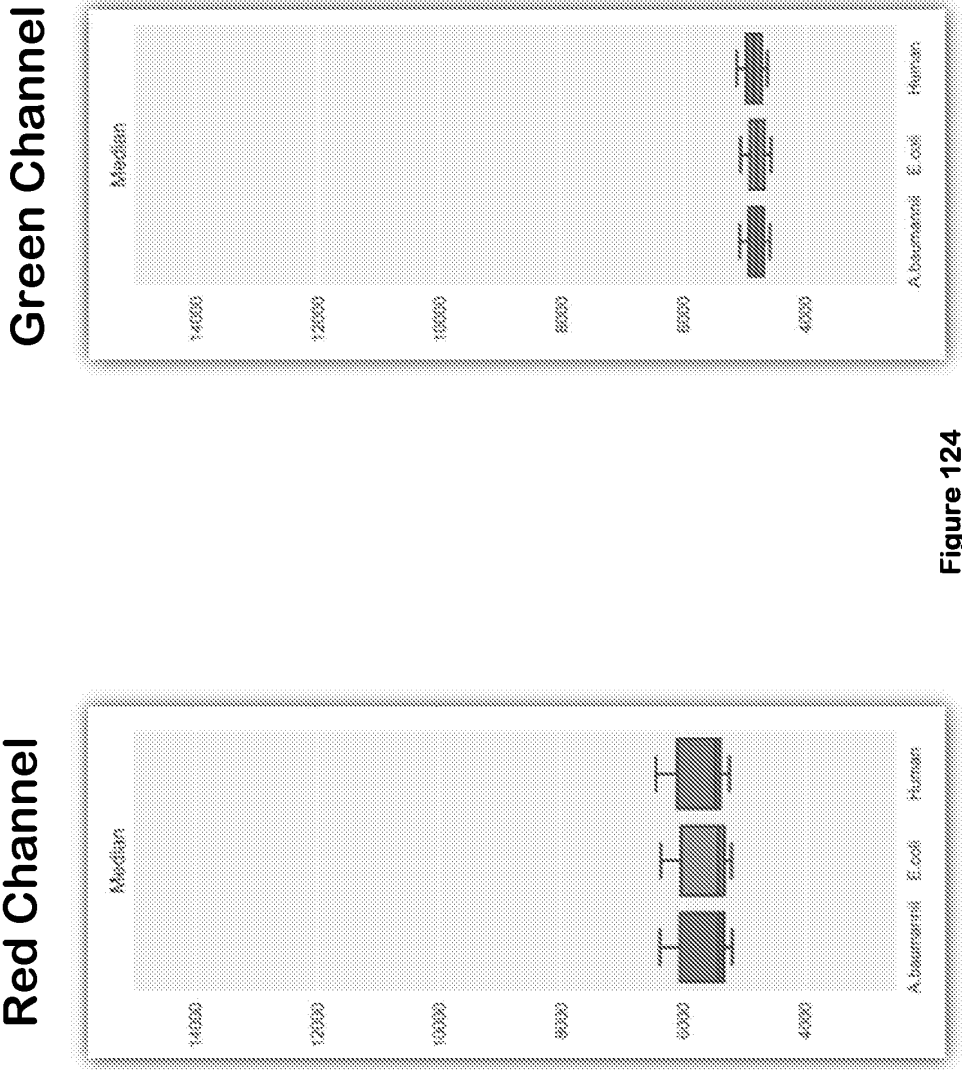
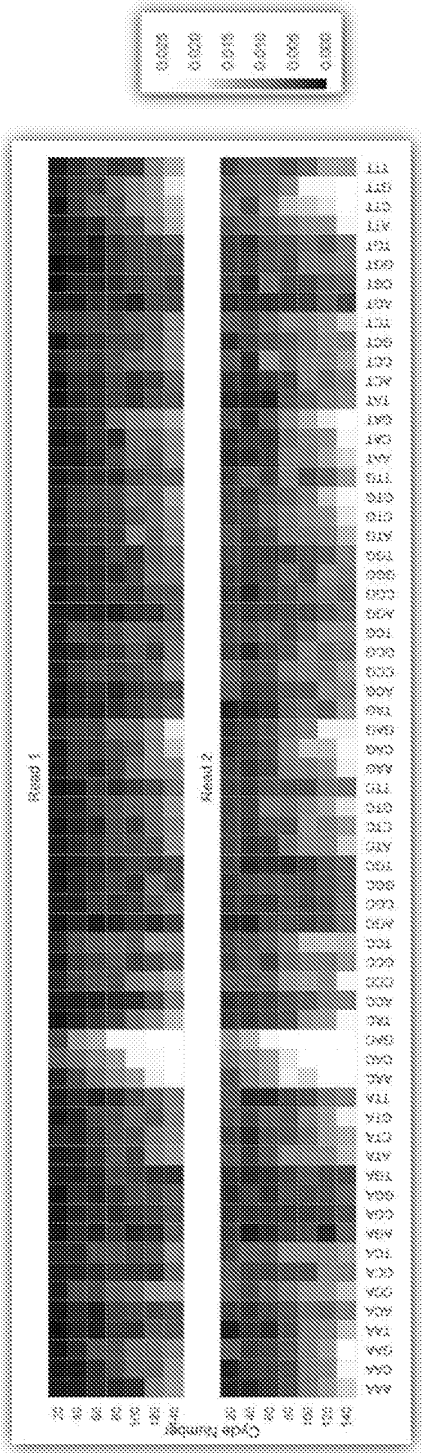


Figure 124

Error Profiles

RTA



Neural Network-Based Base Caller

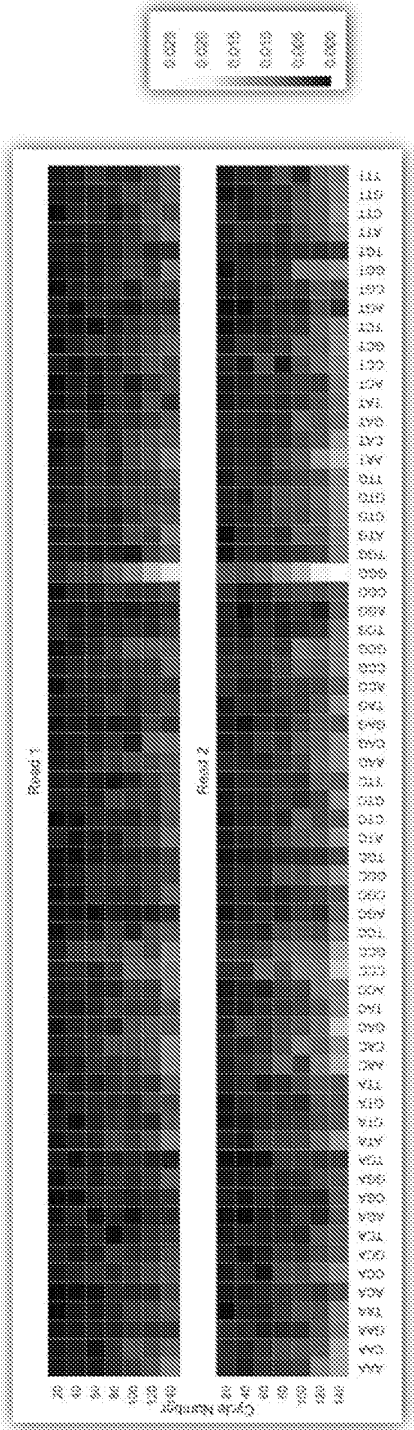


Figure 125

Run to Run Generalization: Four Different Instruments

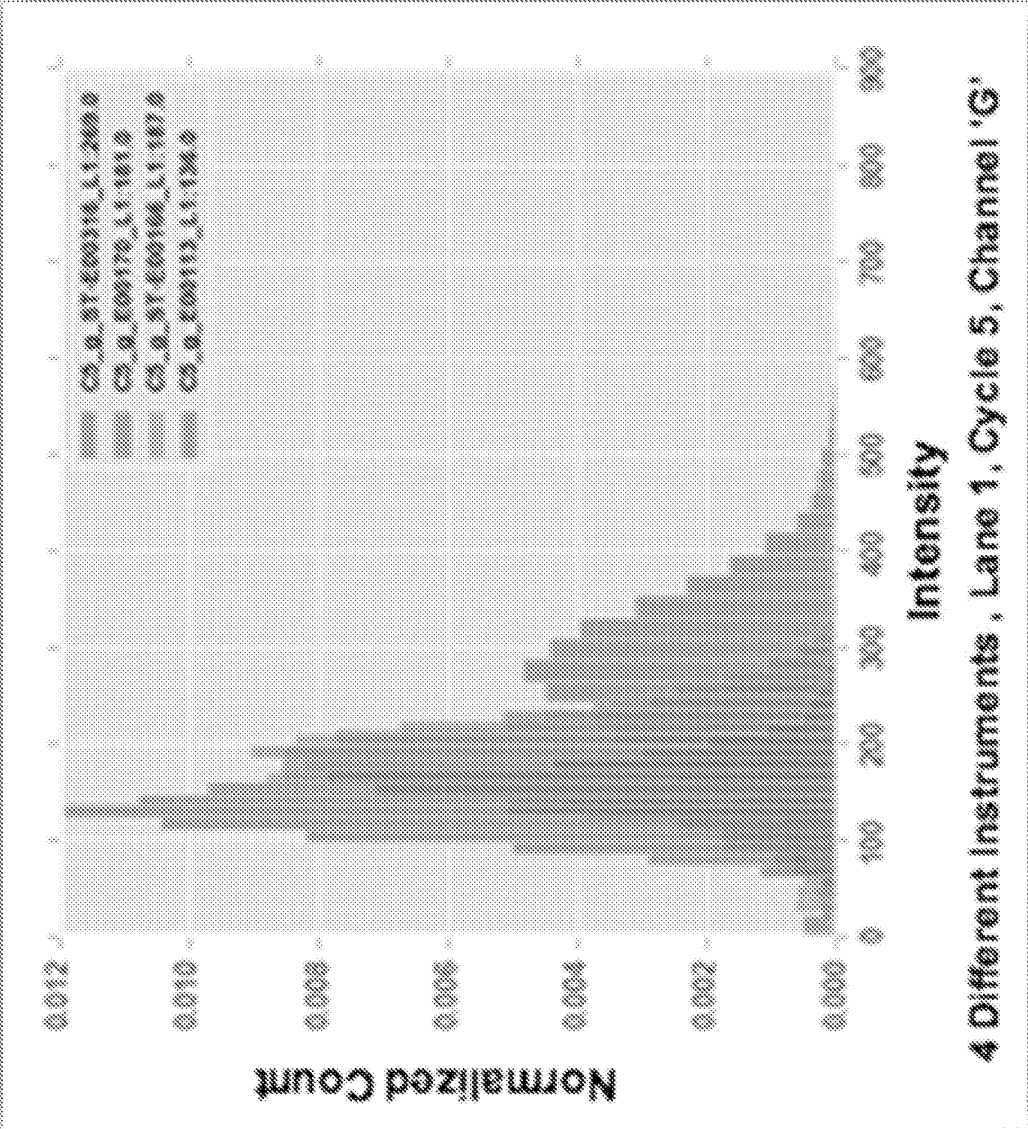


Figure 126a

Run to Run Generalization: Same Instrument for Four Different Runs

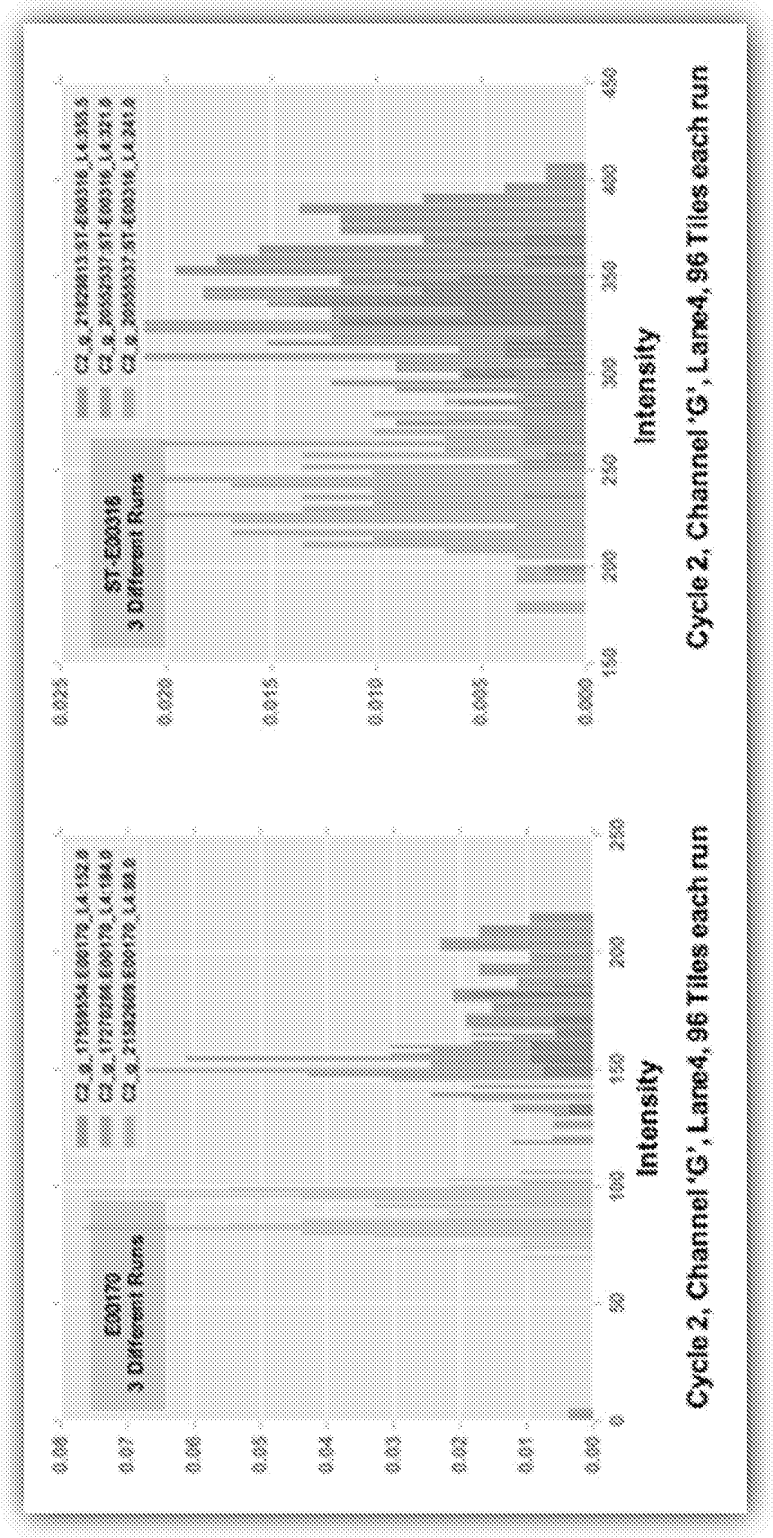


Figure 126b

Genome Statistics

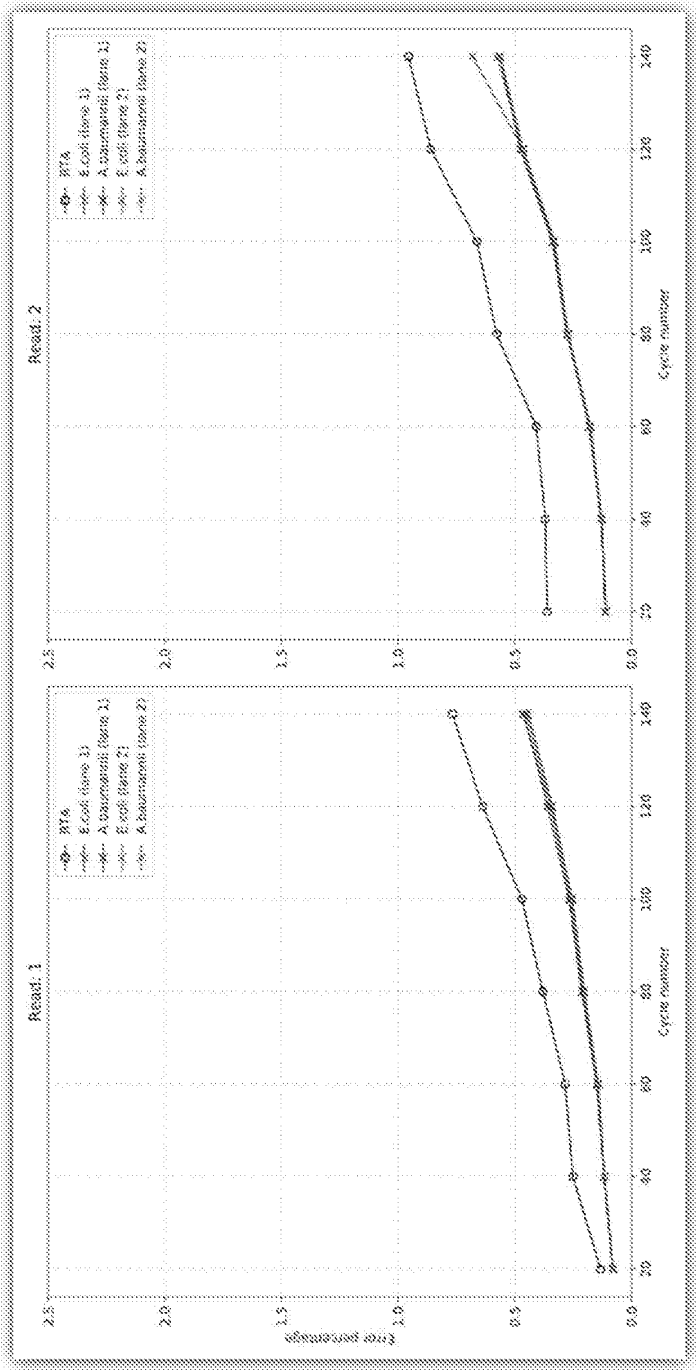


Figure 127

Genome Context

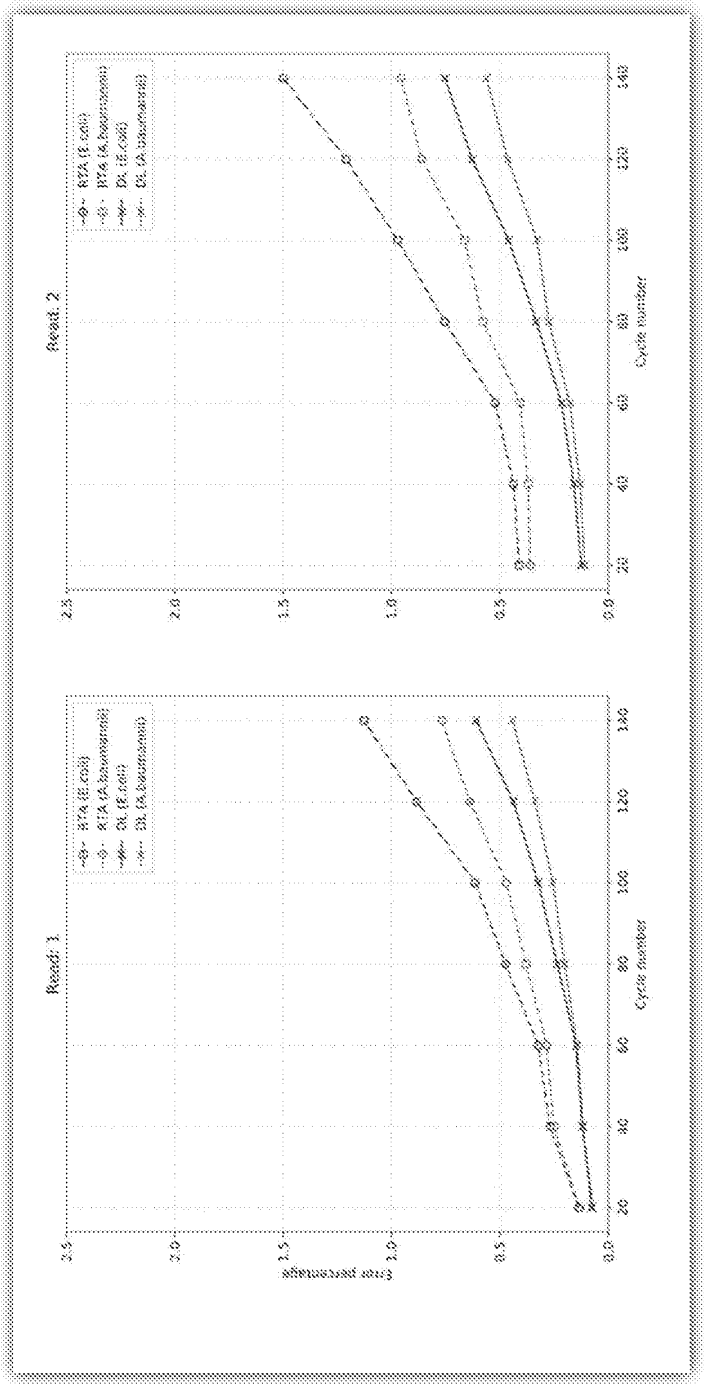


Figure 128

Long Reads

MiniSeq Run:
2x250

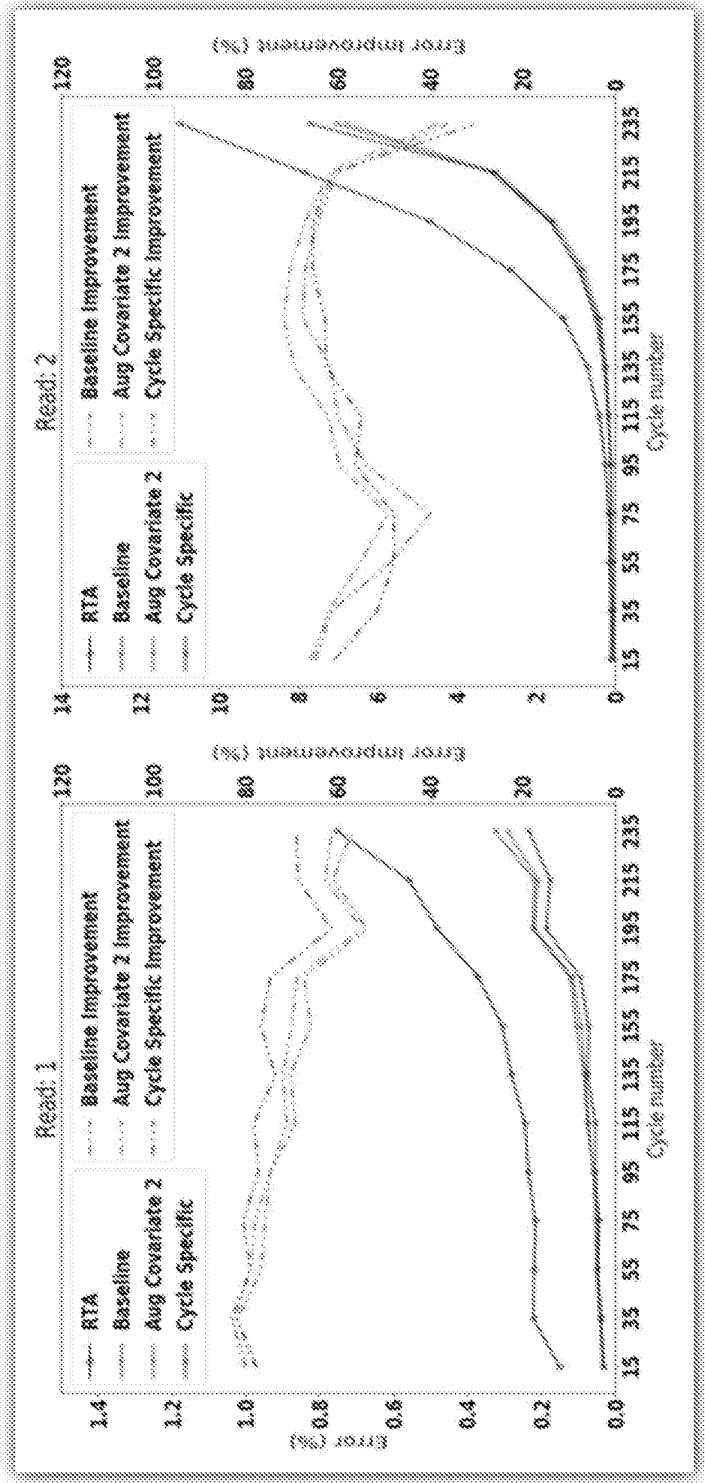


Figure 129

Attention to Central Cluster Pixel(s) and its Neighboring Pixels

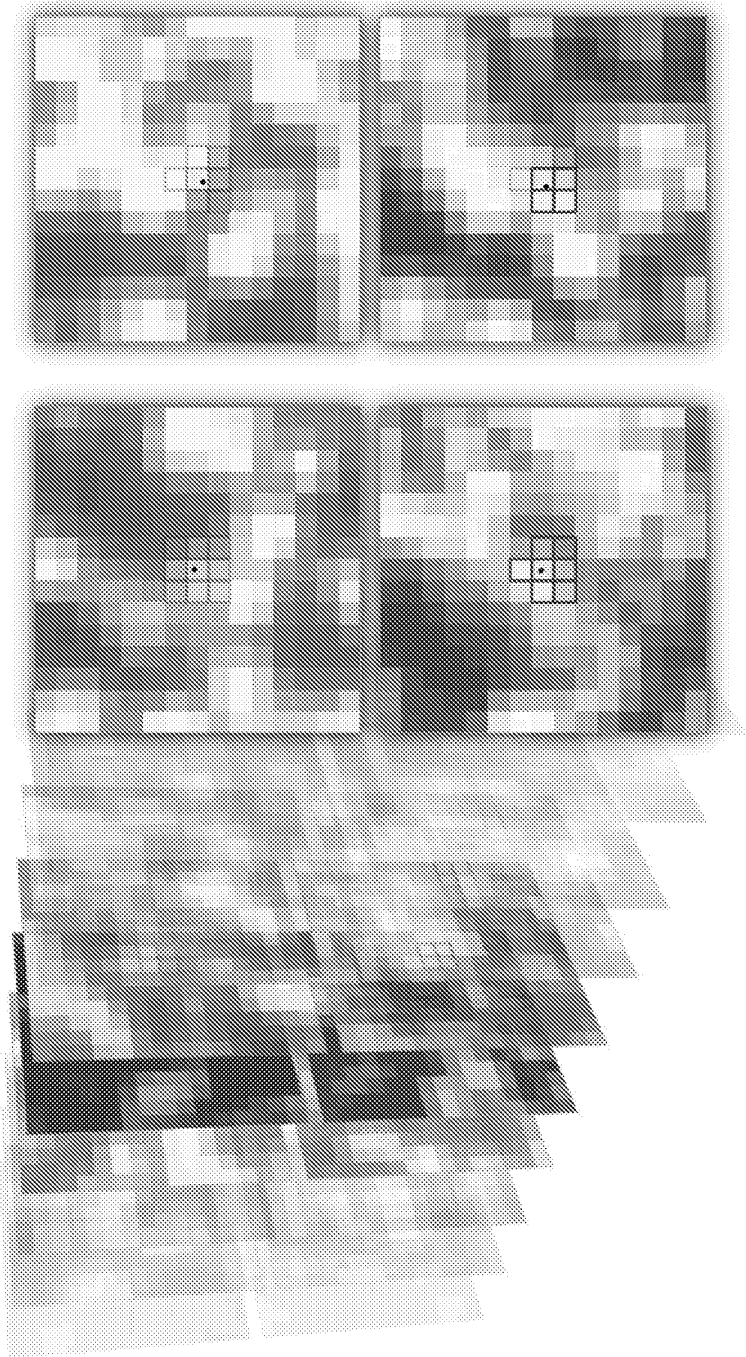


Figure 130

145/169

Hardware

- Inference Engines such as:
 - Jetson (Nvidia)
 - Mythic
- Software based enhancements:
 - Works with all platforms including FPGA
 - Collaboration with UCSD
 - Available FPGA capacity on instrument
- Cloud based solutions
 - Firefly / Dragonfly
- Memory

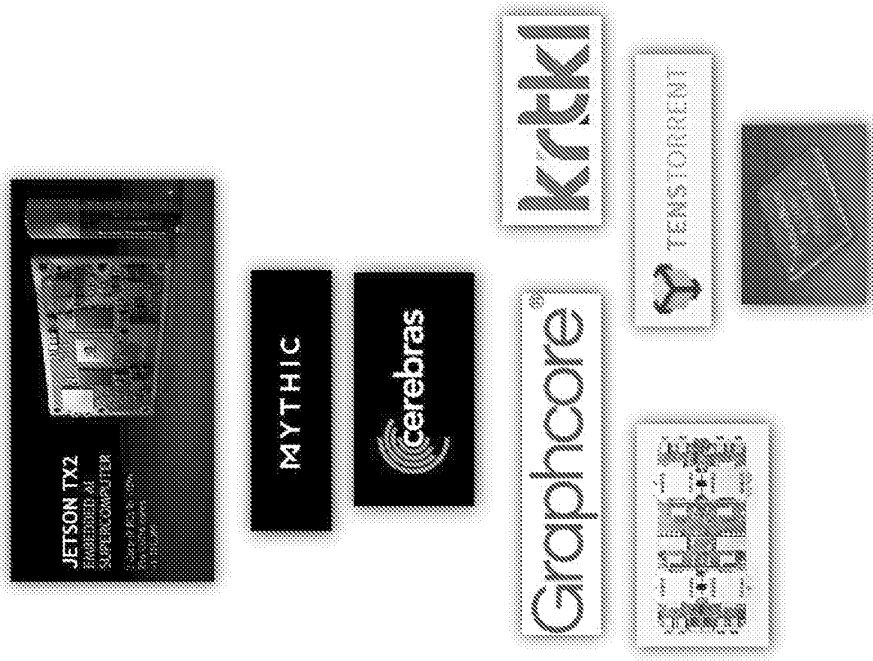


Figure 131

Base Calling

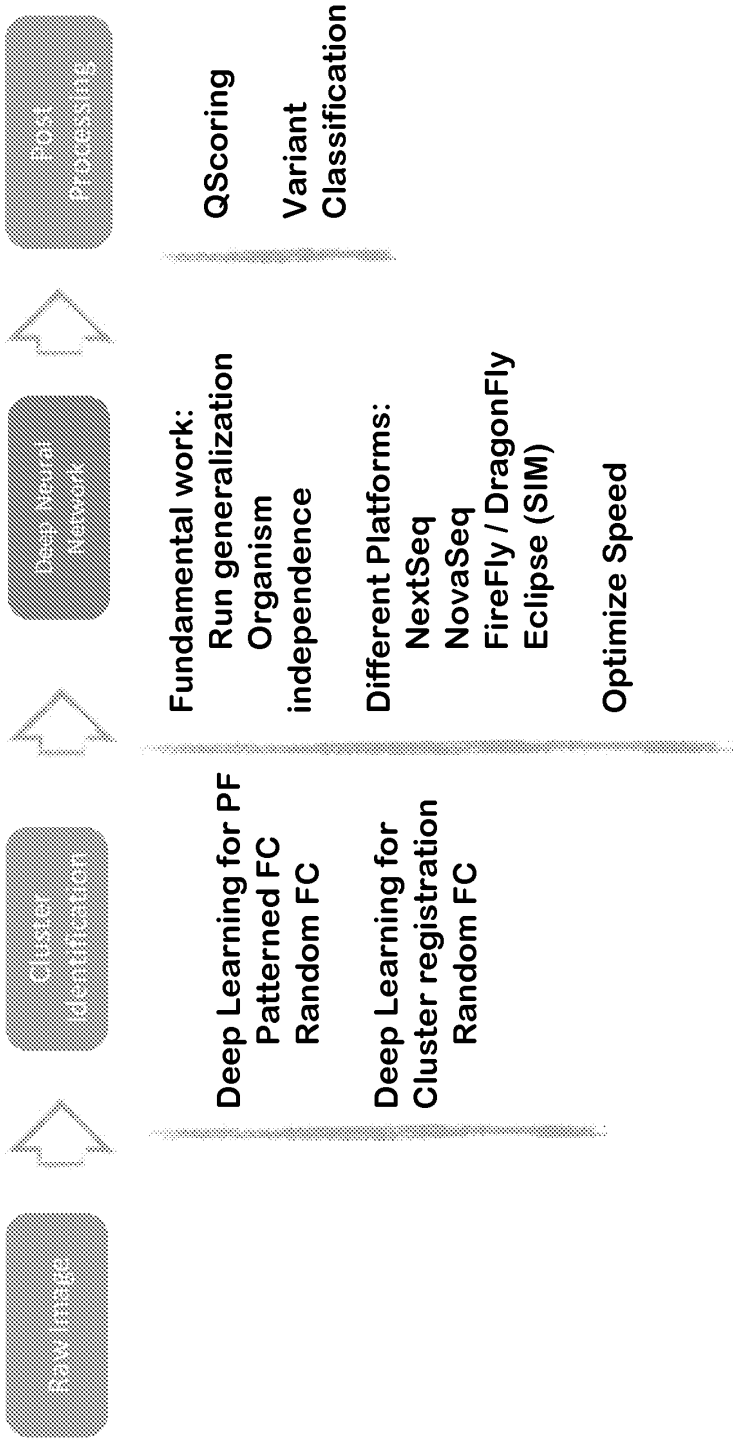


Figure 132

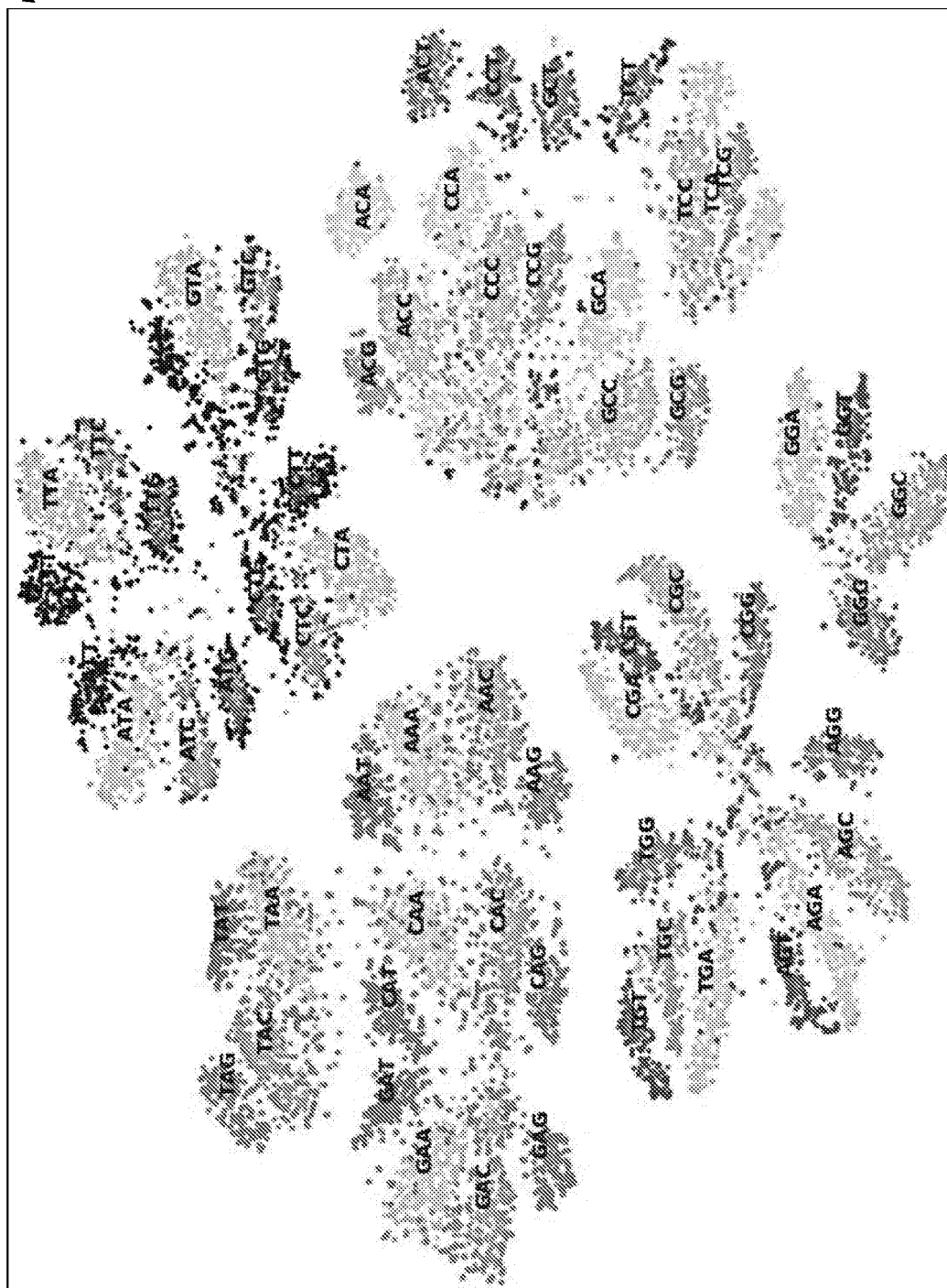
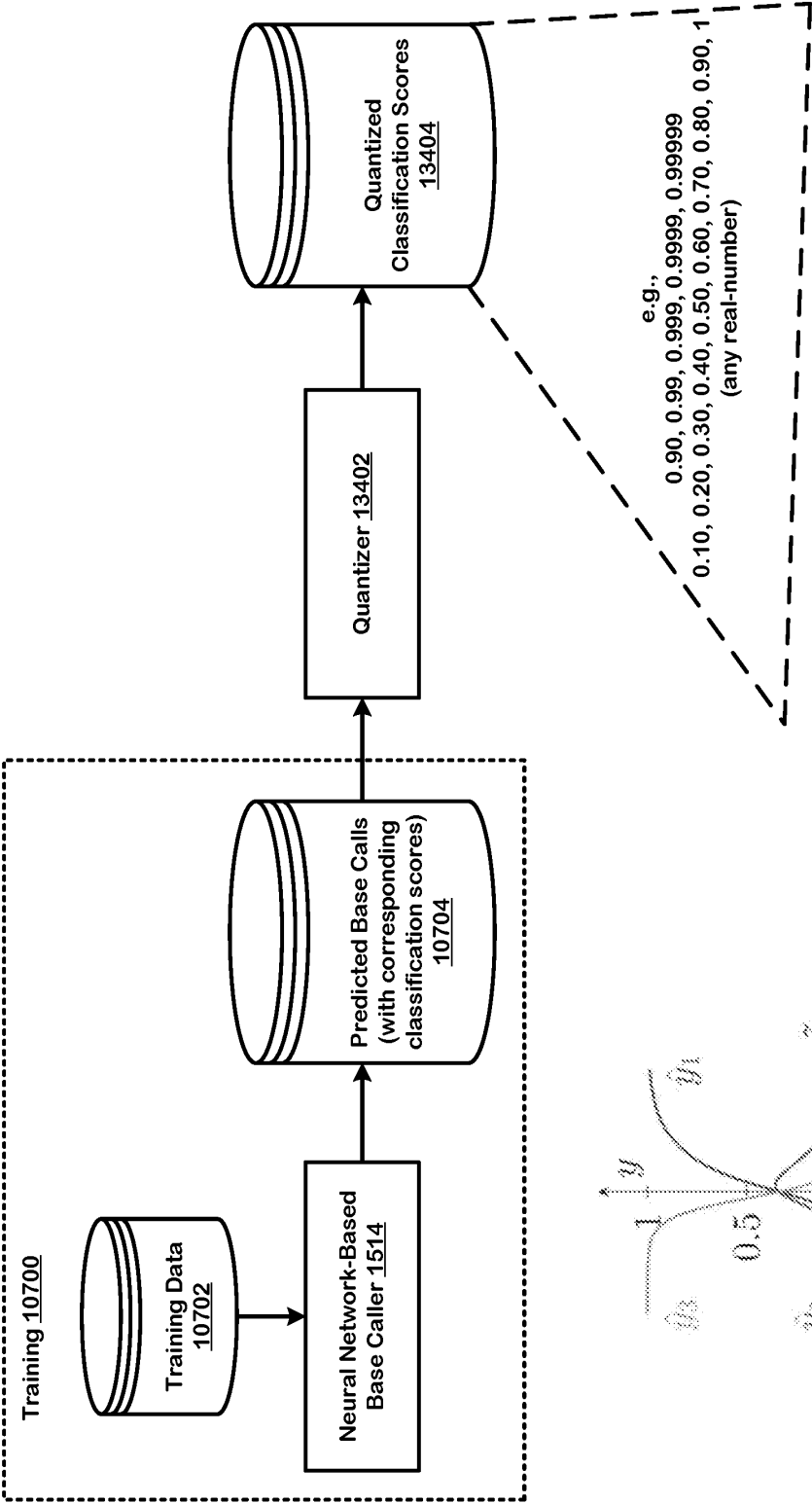


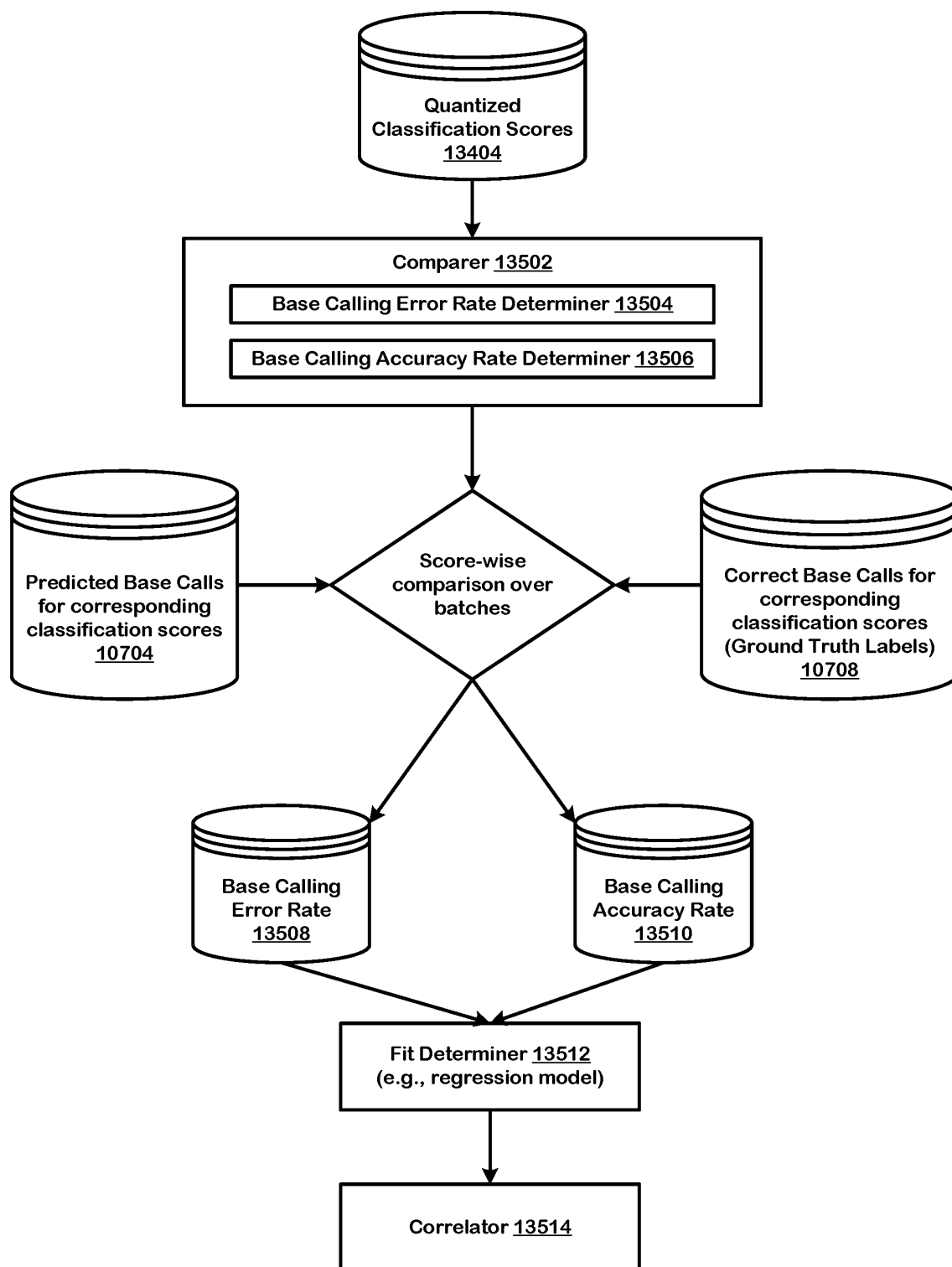
Figure 133

Selecting 13400 Base Call Confidence Probabilities of the Neural Network-Based Base Caller 1514



Softmax 13406

Figure 134

Neural Network-Based Quality Scoring 13500**Figure 135**

Quality Score Correspondence 13600

Figure 136a

Quality Score Correspondence Scheme <u>13600a</u>	
Quality Score	Softmax Score
Q10	0.90
Q20	0.99
Q30	0.999
Q40	0.9999
Q50	0.99999
Q60	0.999999

Quality Score Correspondence Scheme <u>13600b</u>	
Binned Quality Score	Softmax Score
Q2-Q9 -> Q06	0.80
Q10-Q19 -> Q15	0.95
Q20-Q24 -> Q22	0.993
Q25-Q29 -> Q27	0.997
Q30-Q34 -> Q33	0.9991
Q35-Q39 -> Q37	0.9995
Q40 or > Q40 -> Q40	0.9999

Figure 136b

Quality Scoring during Inference 13700

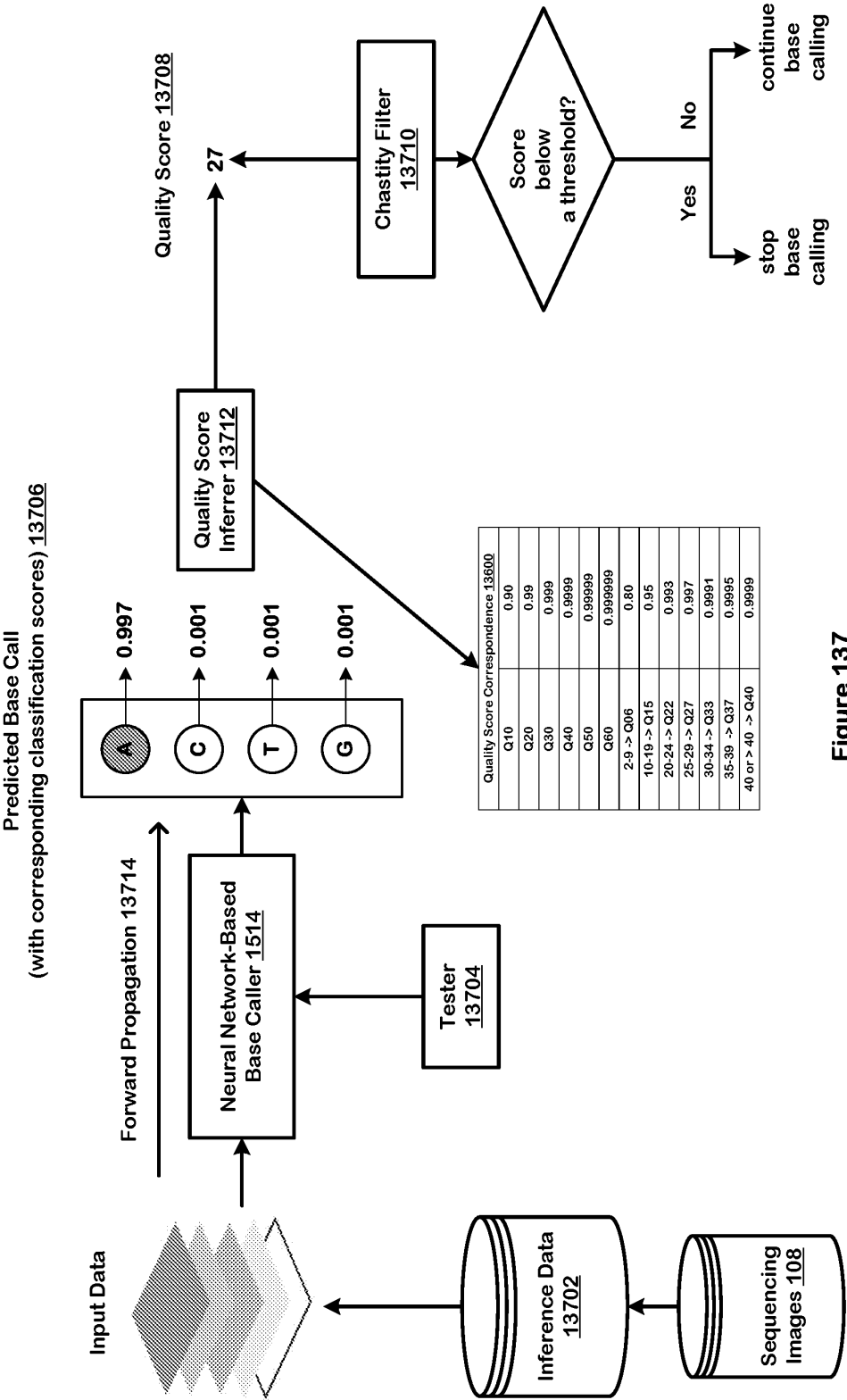


Figure 137

Training 13800 the Neural Network-Based Quality Scorer 13802

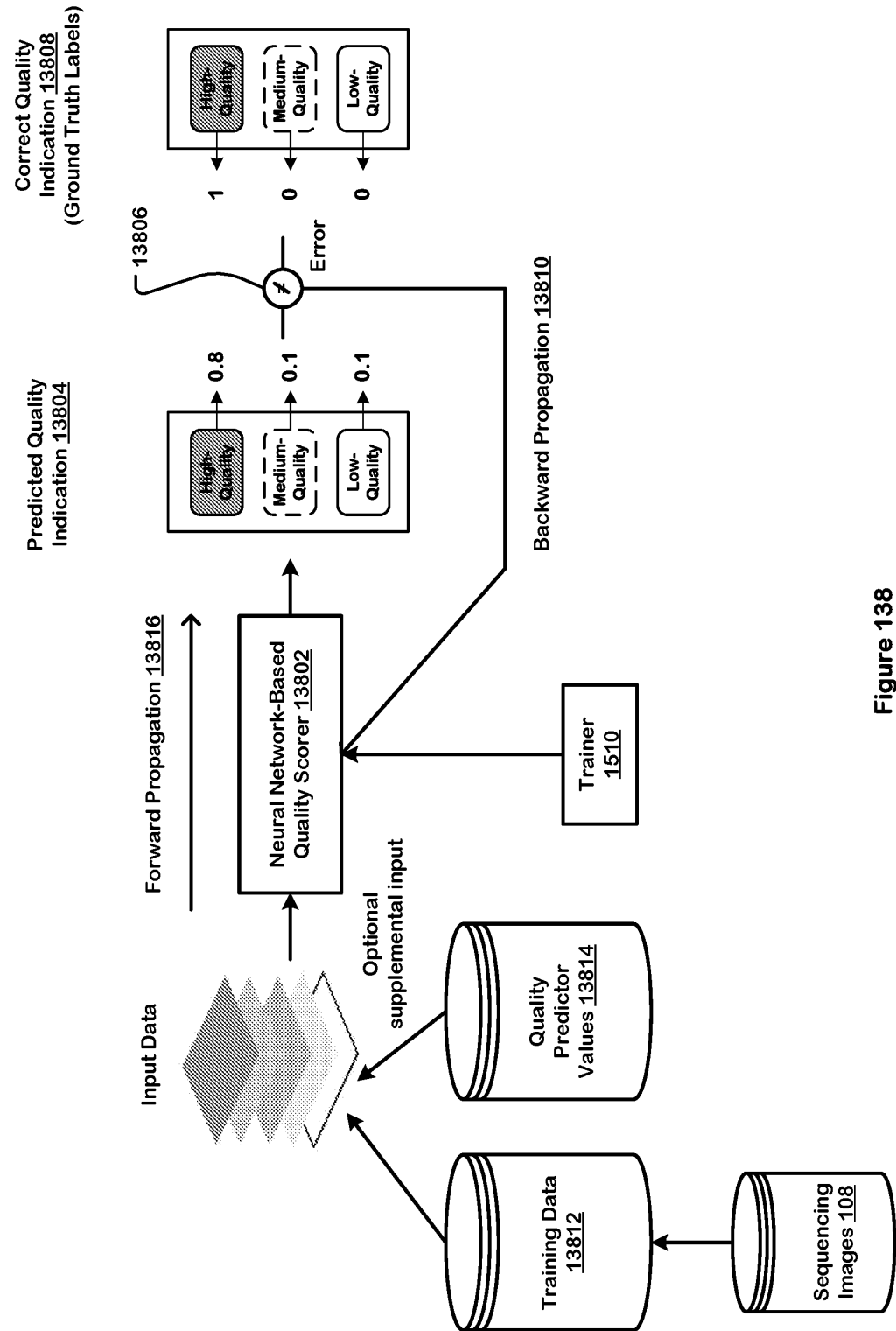


Figure 138

Inference 13900 with the Neural Network-Based Quality Scorer 13802

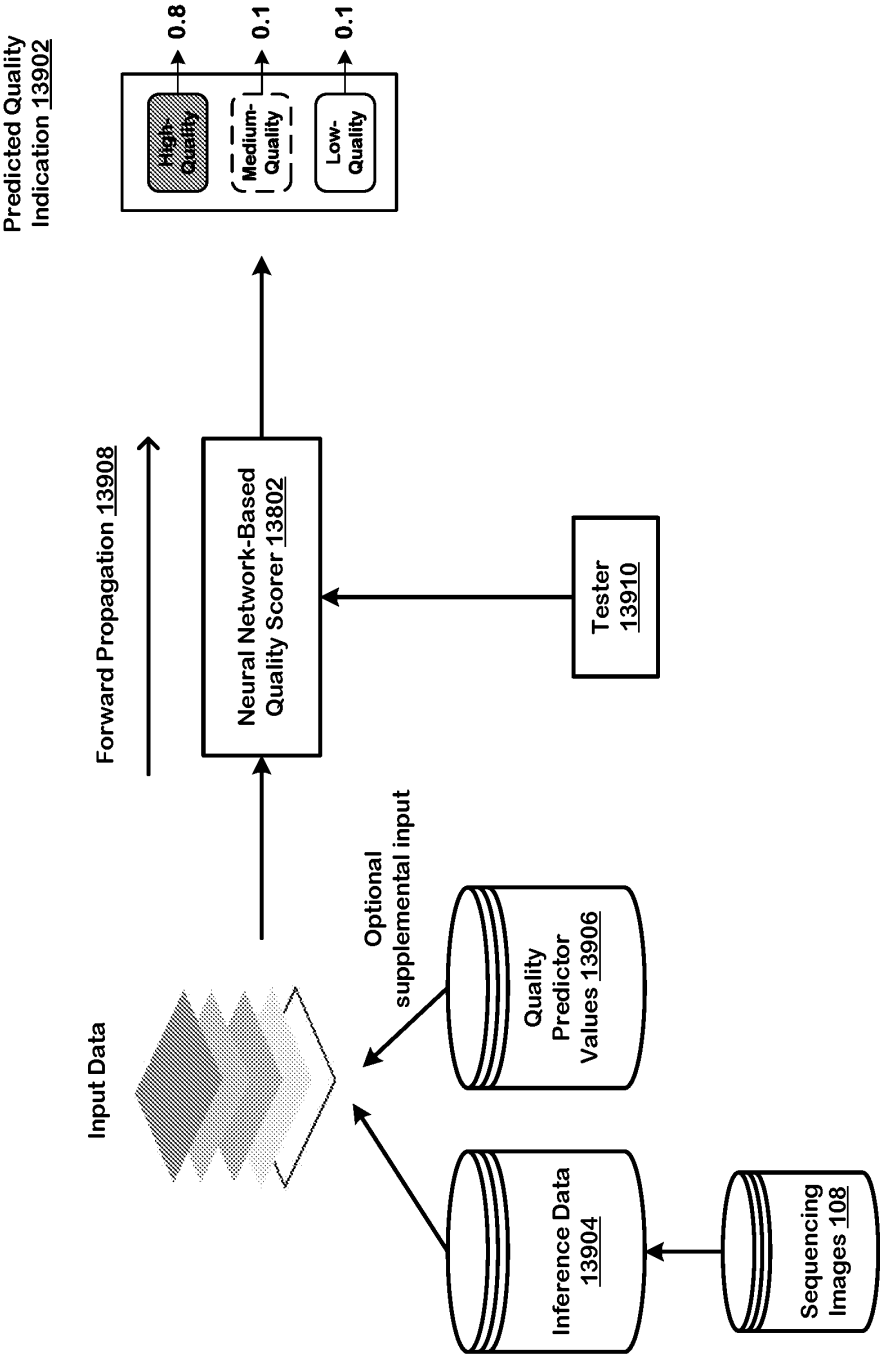


Figure 139

Lossless Transformation 14000

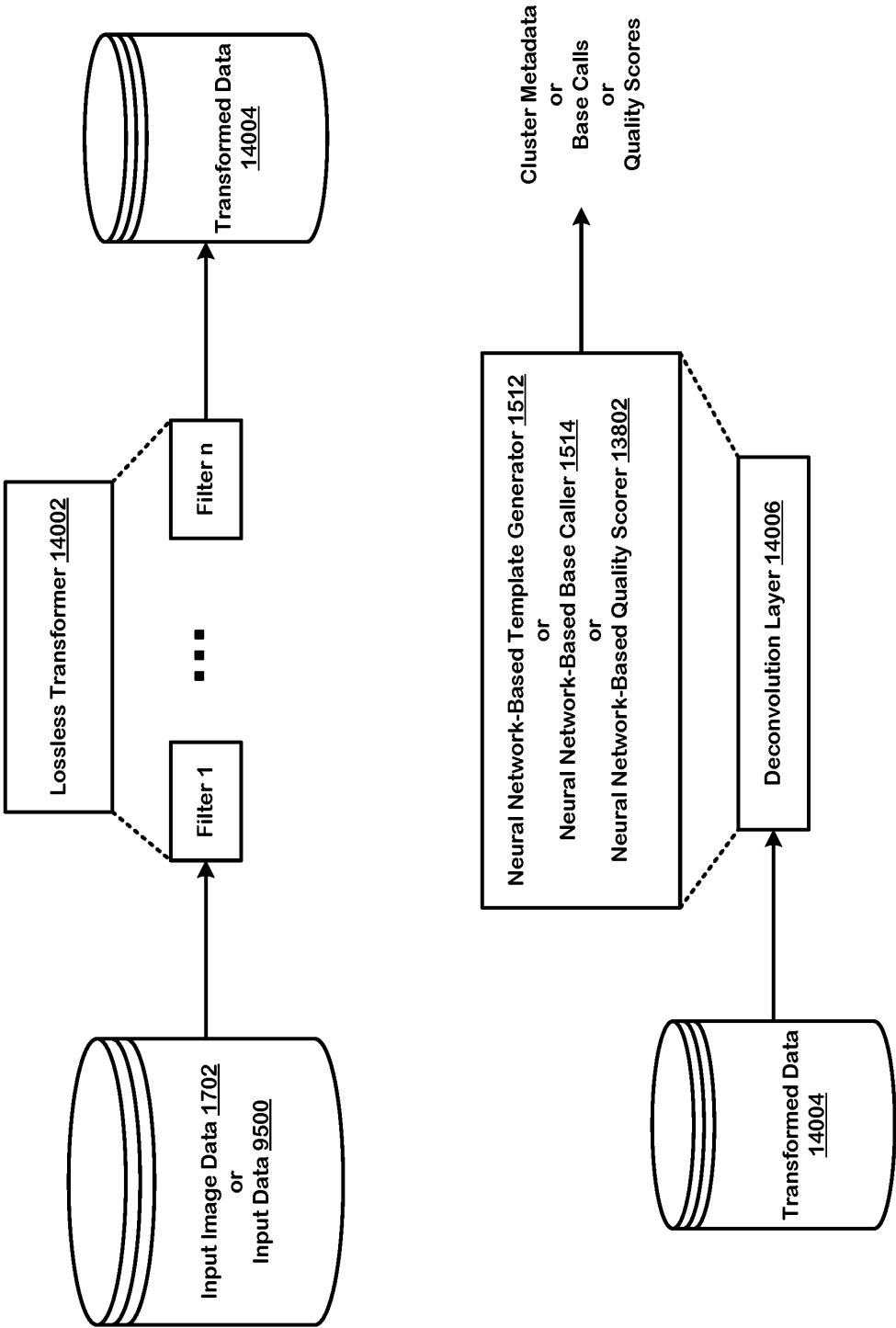


Figure 140

155/169

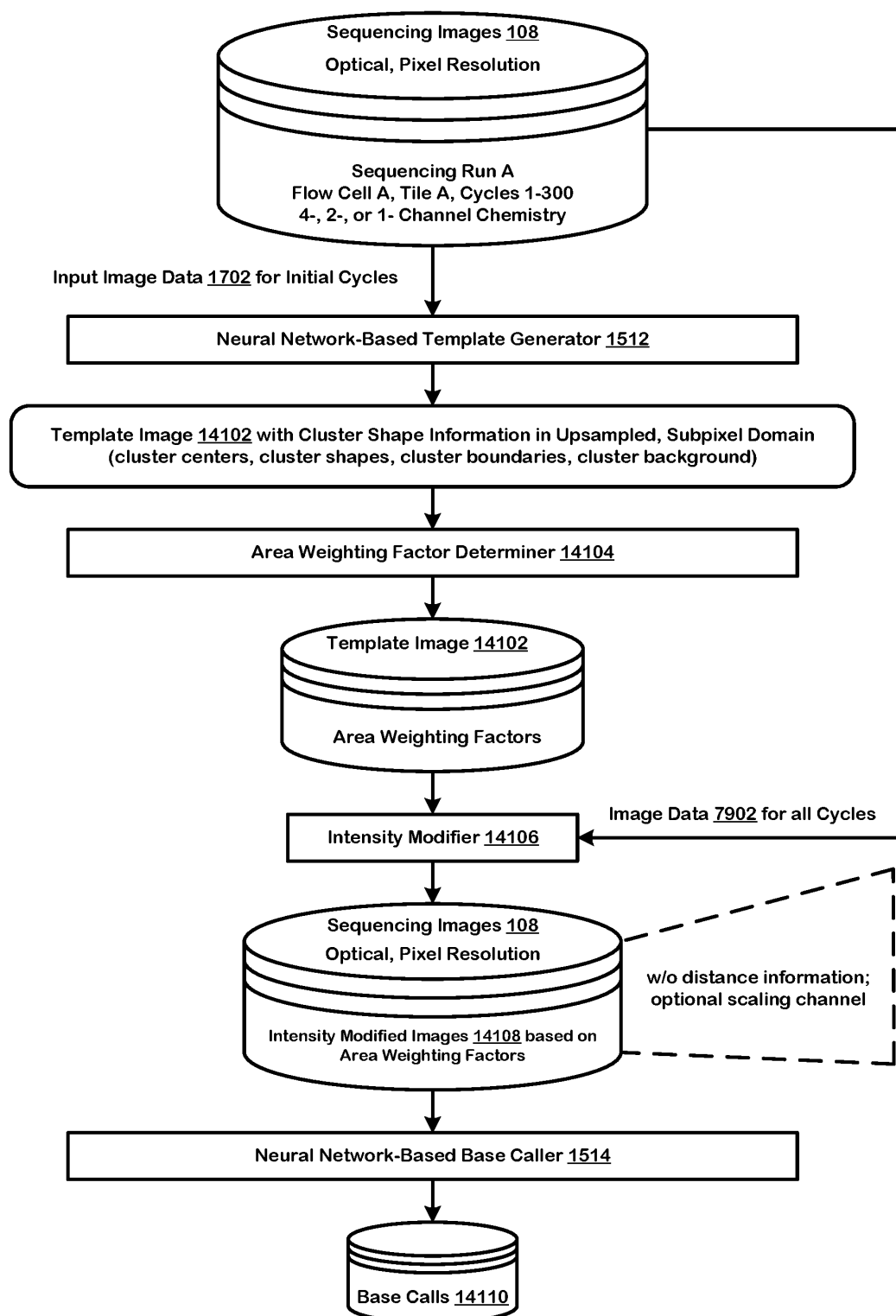


Figure 141

156/169

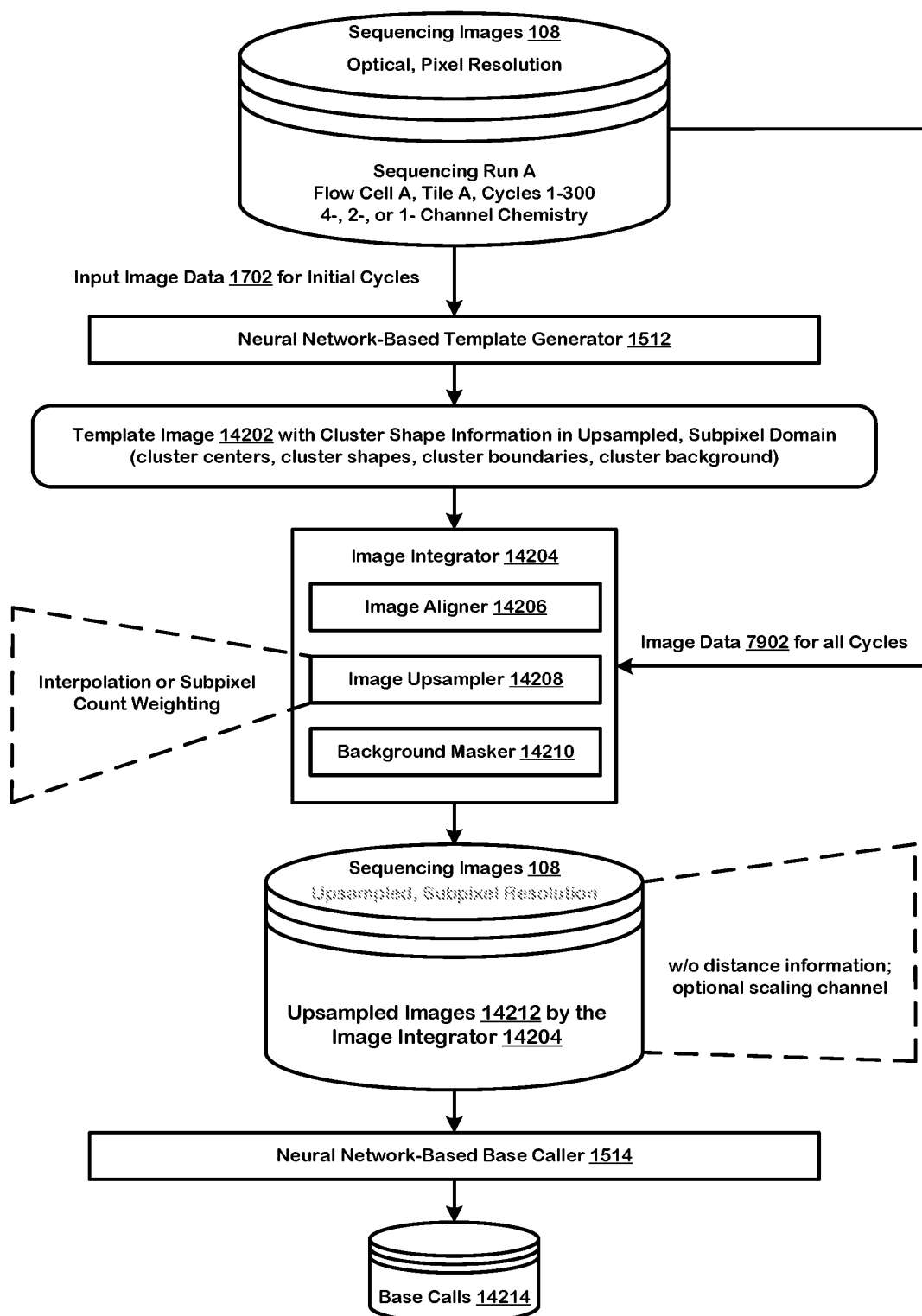
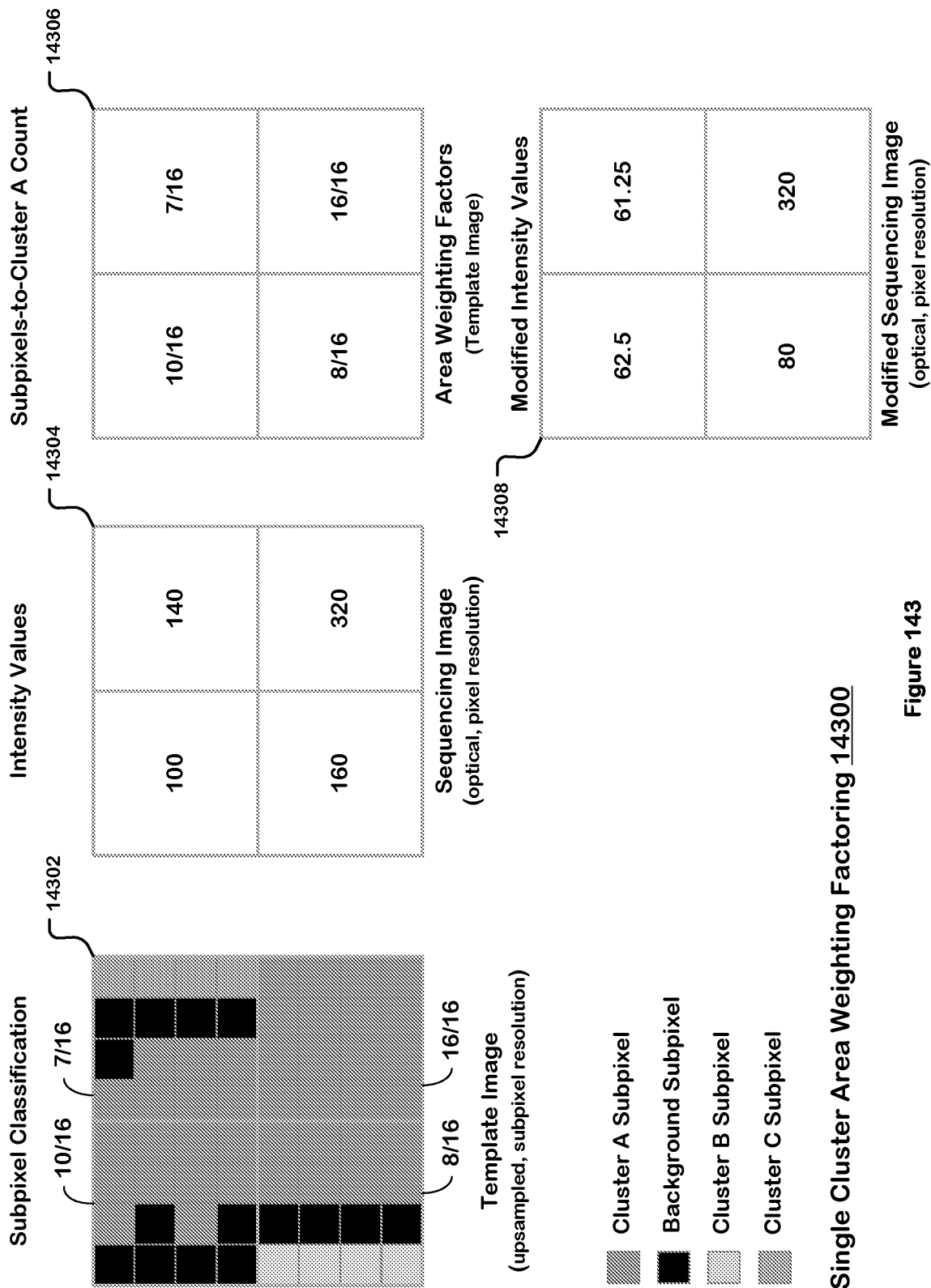
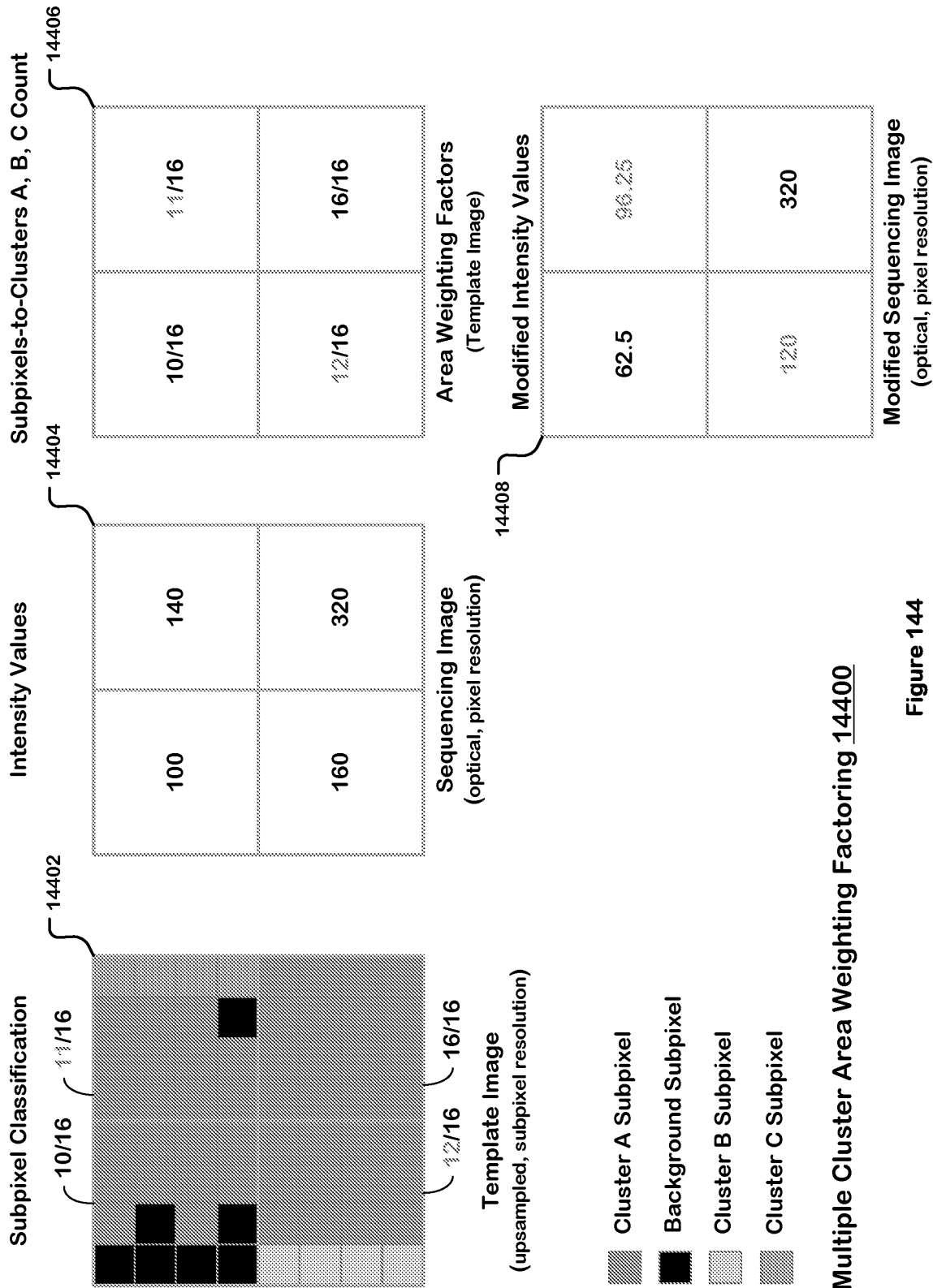
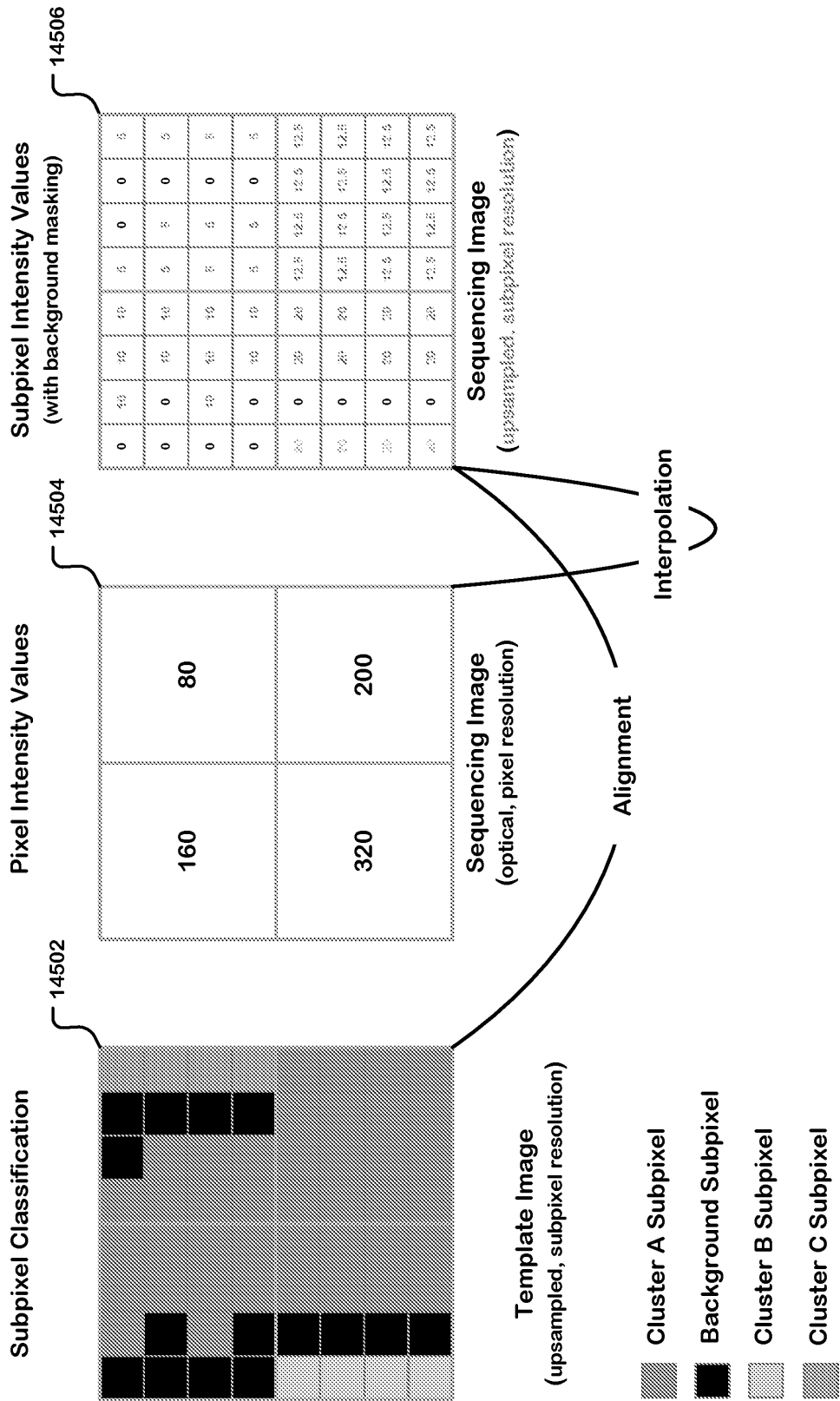


Figure 142

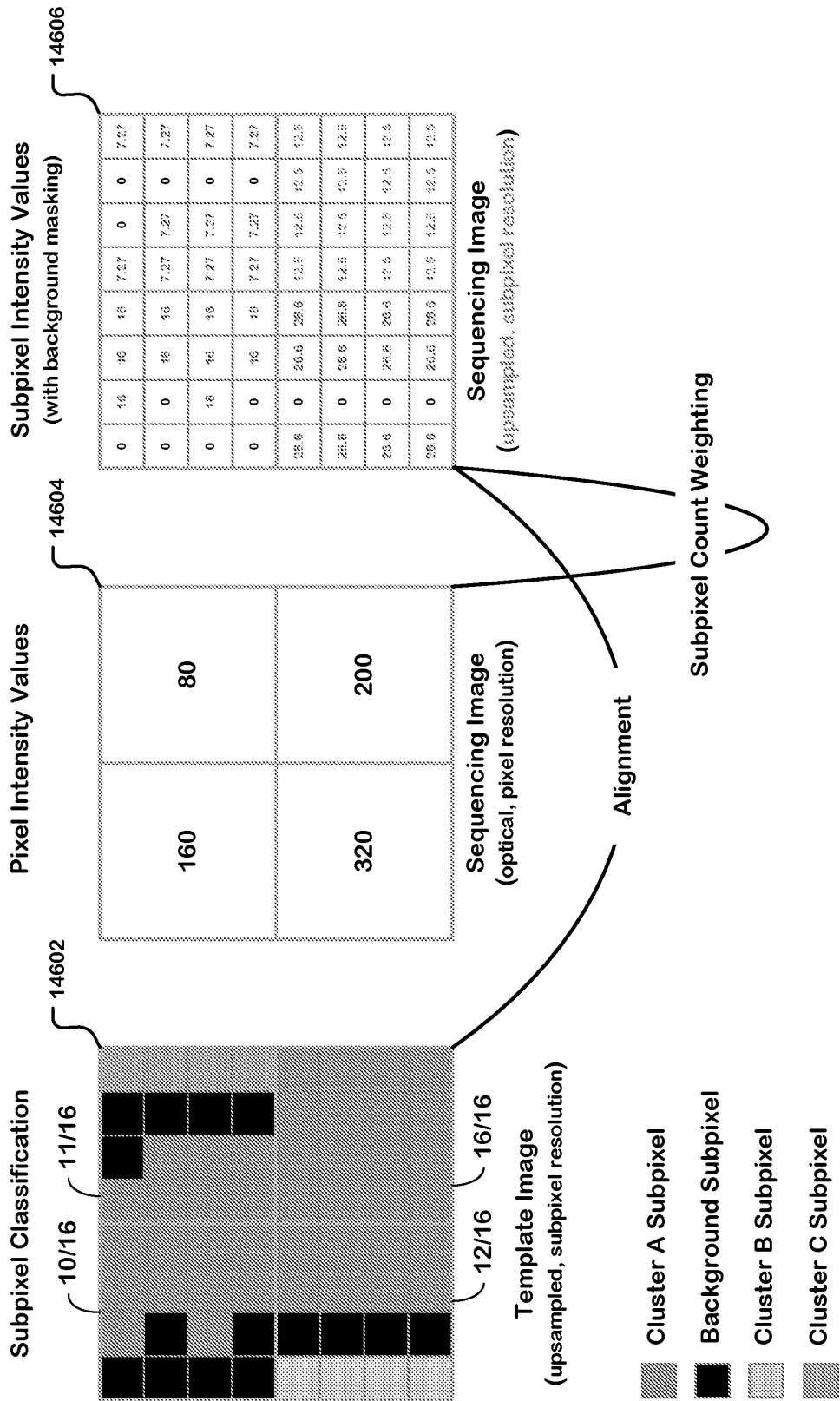






Upsampling & Background Masking 14500

Figure 145



Upsampling & Background Masking 14600

Figure 146

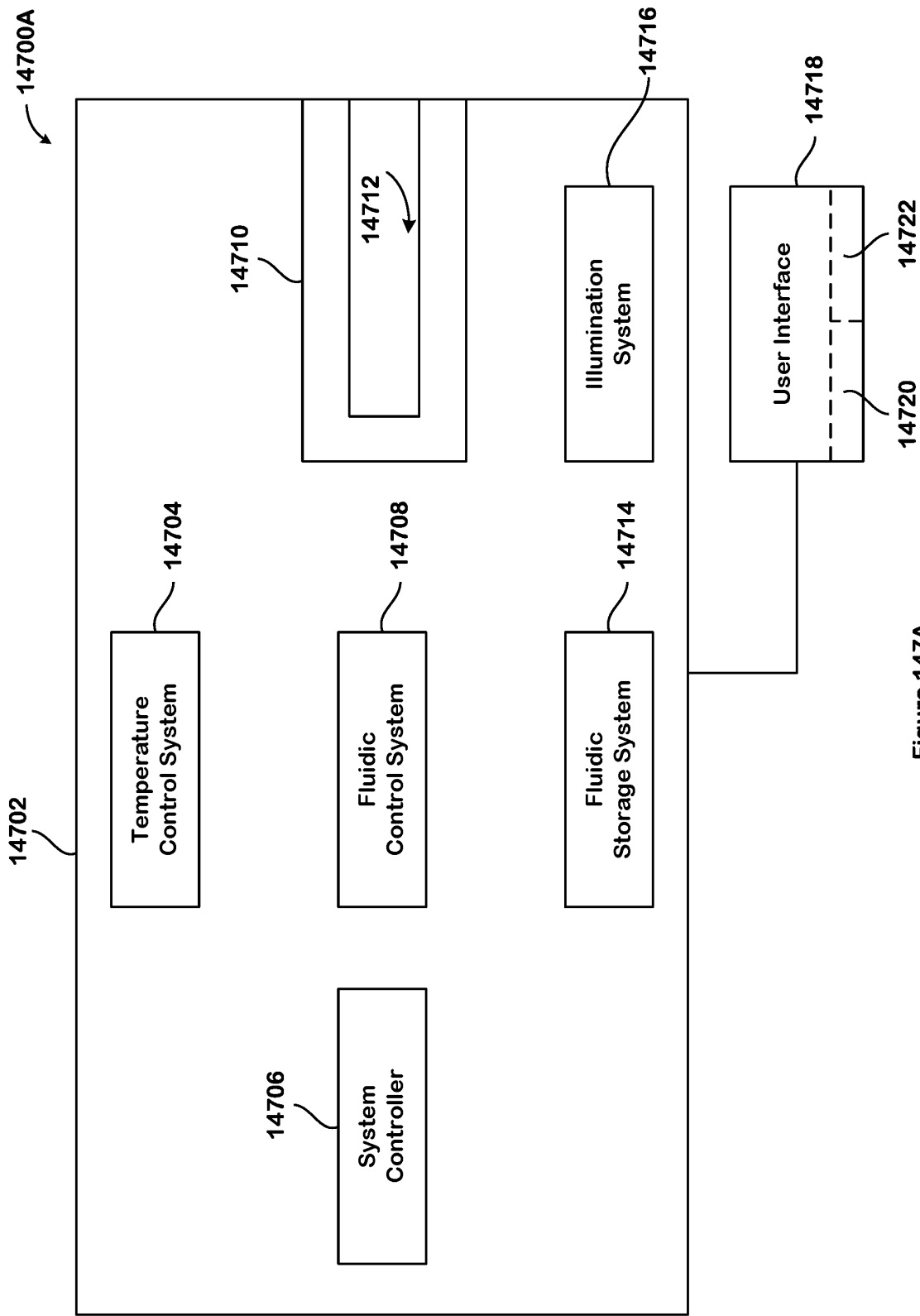


Figure 147A

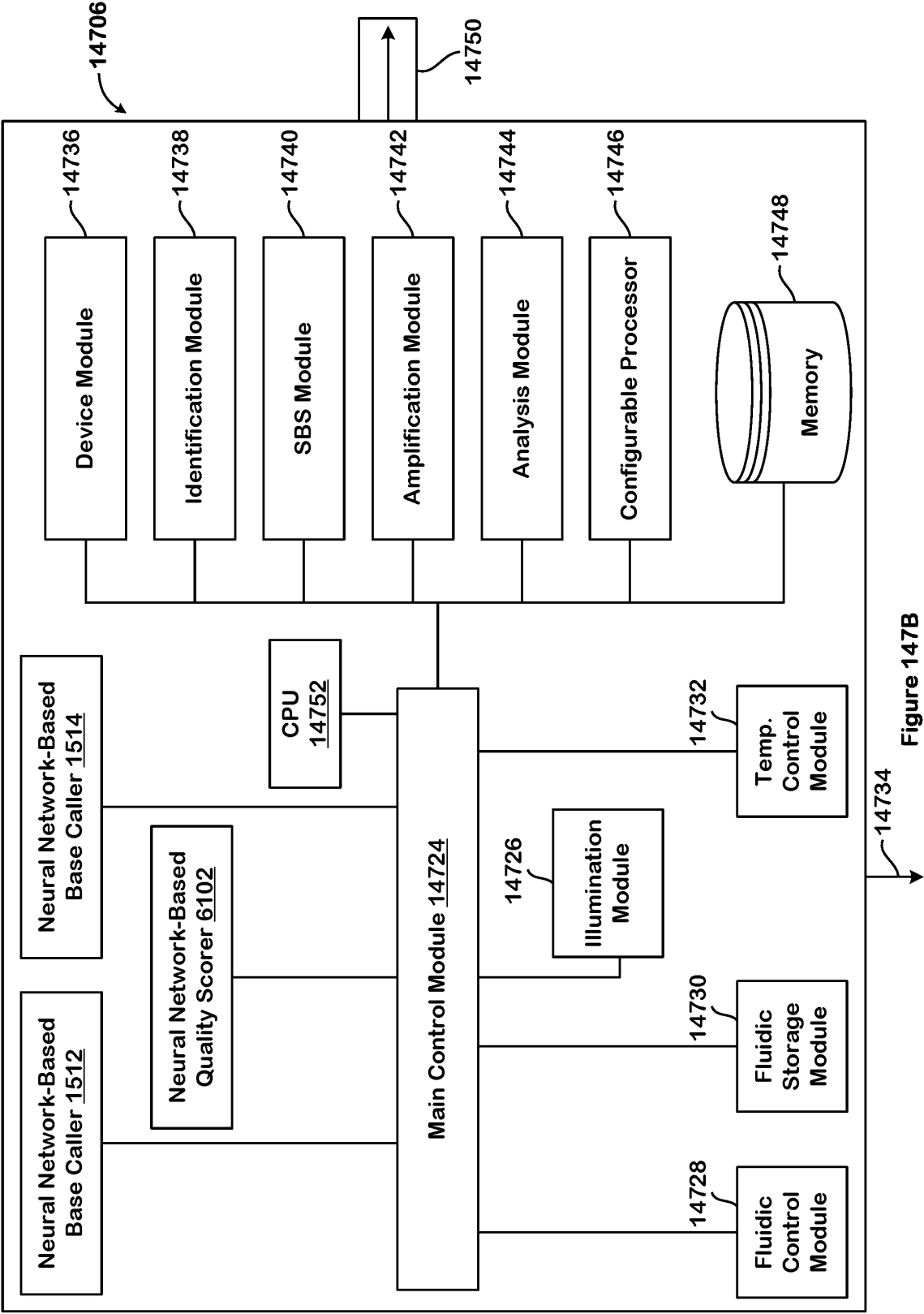


Figure 147B

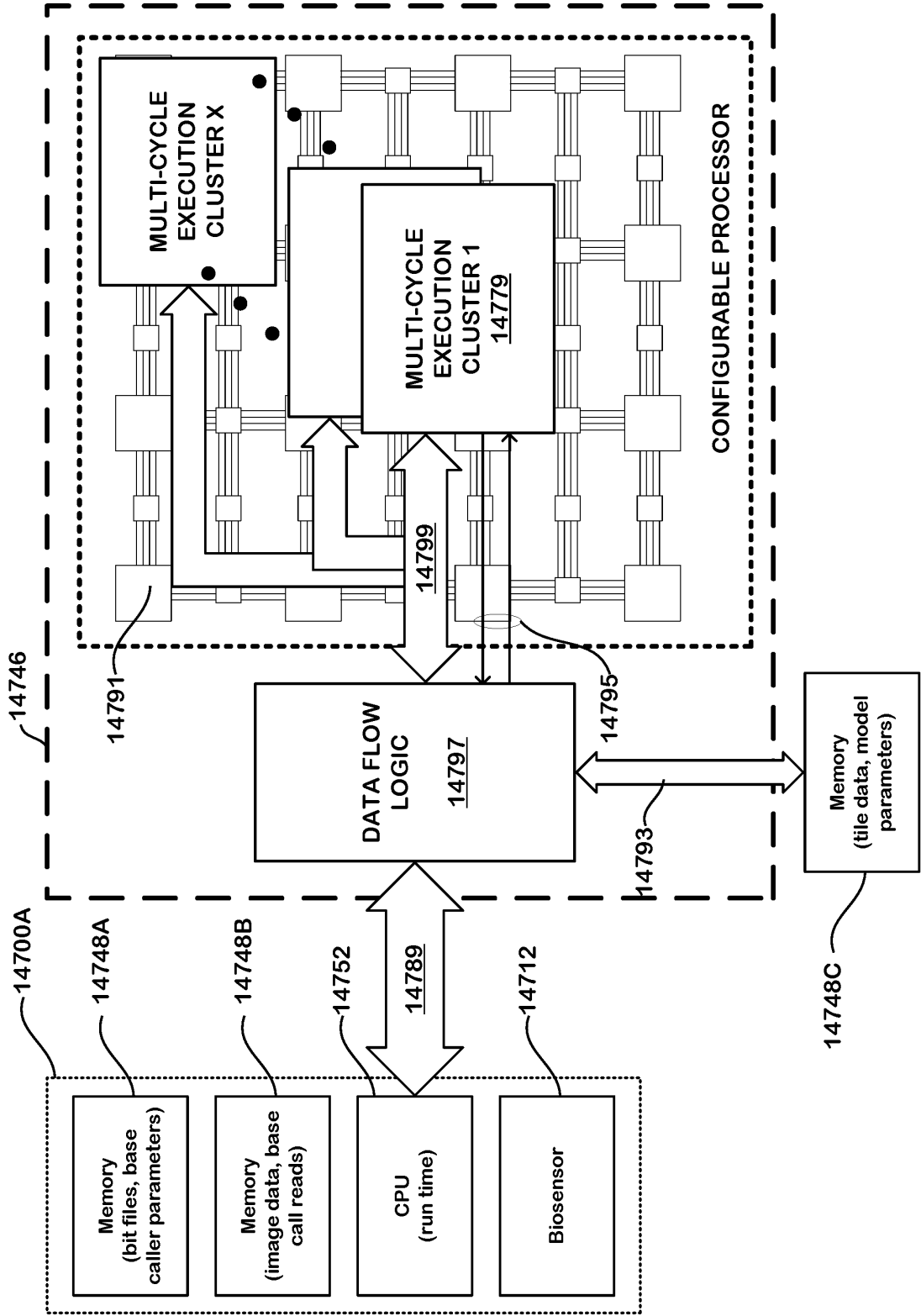


Figure 147C

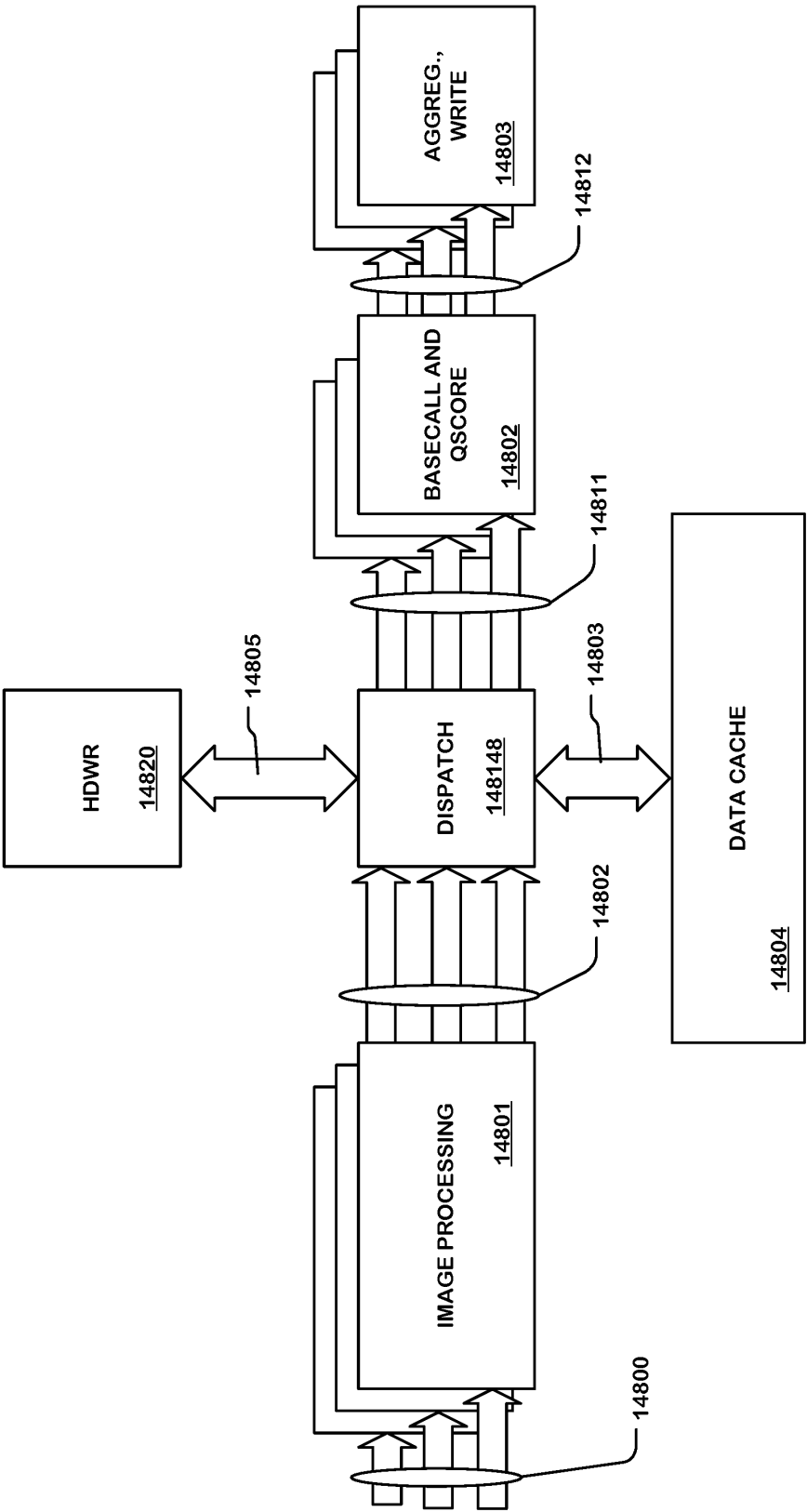


Figure 148A

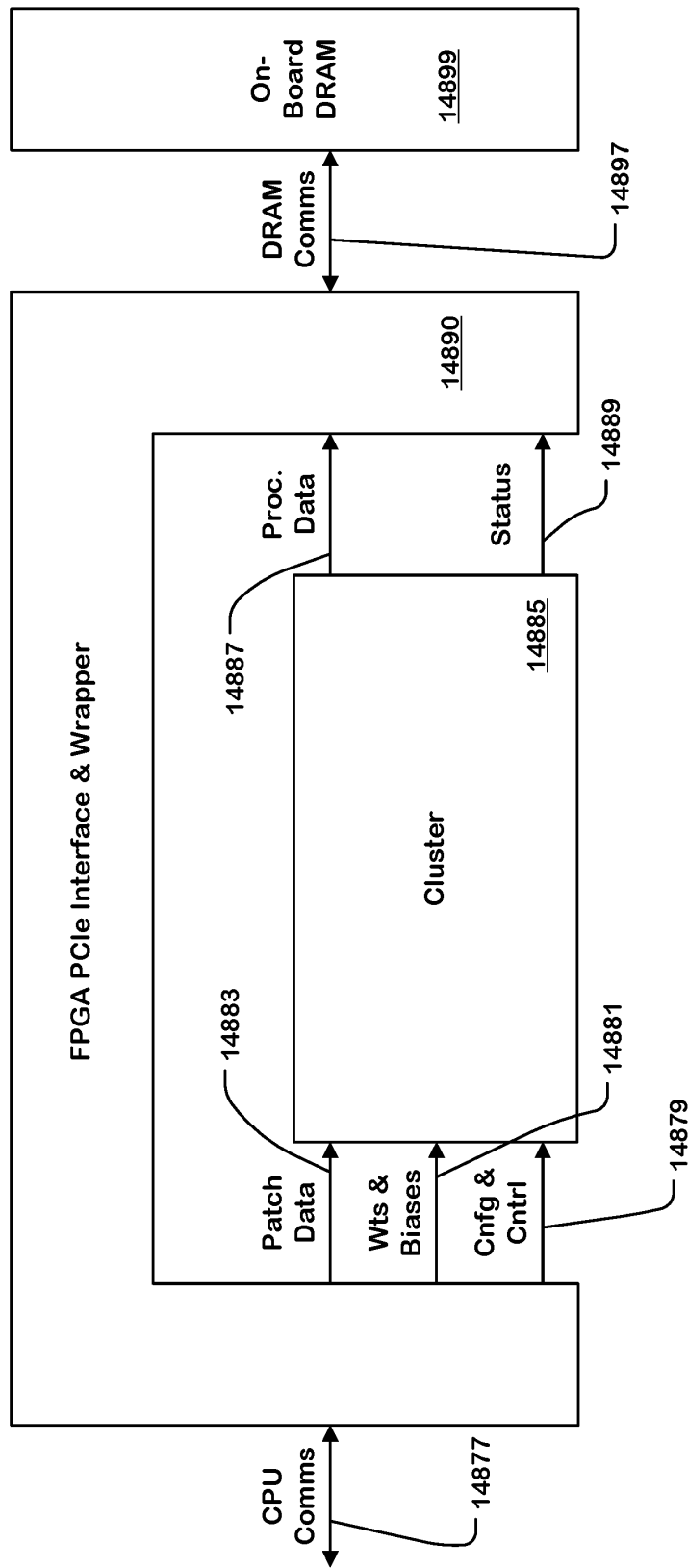


Figure 148B

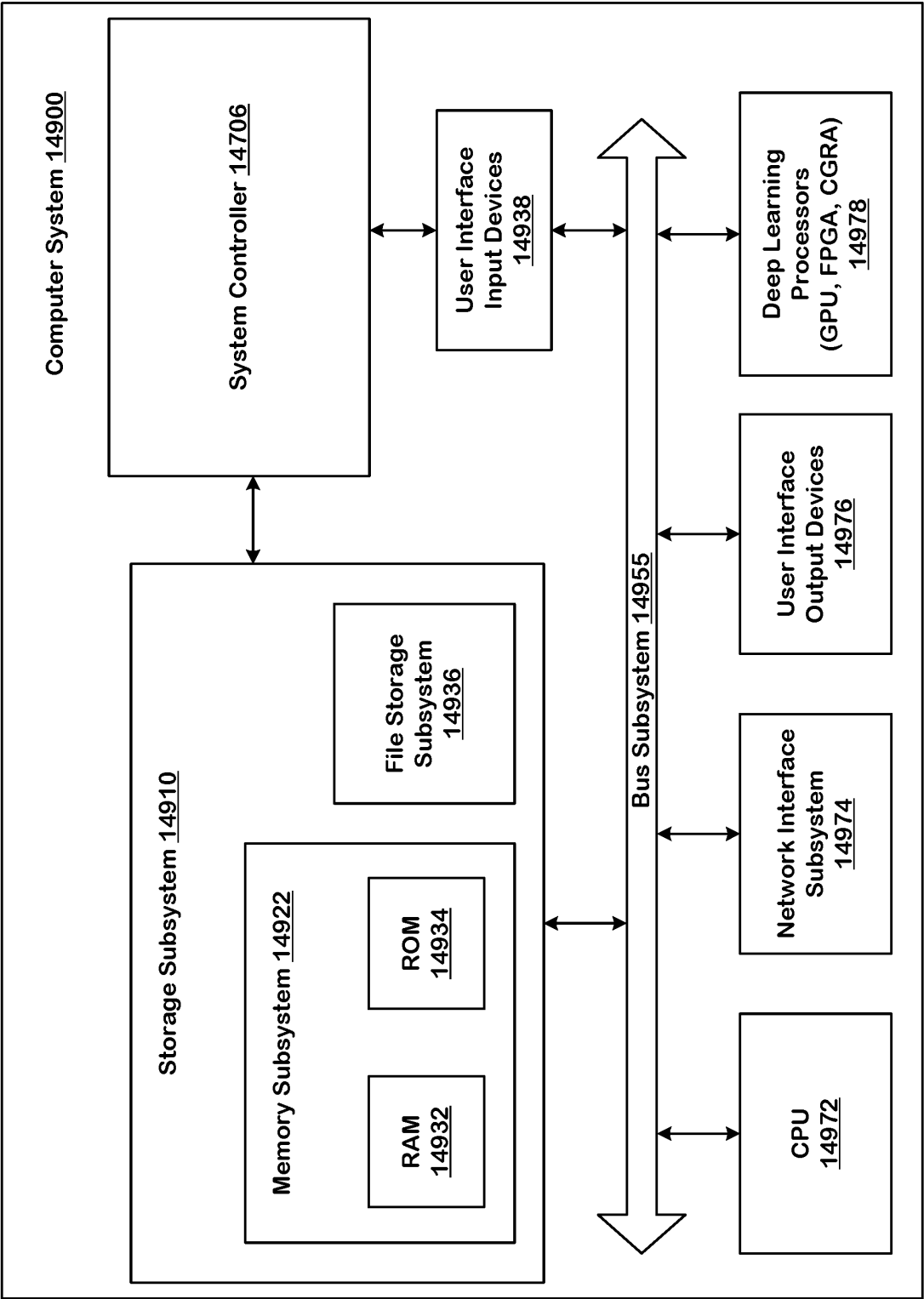


Figure 149

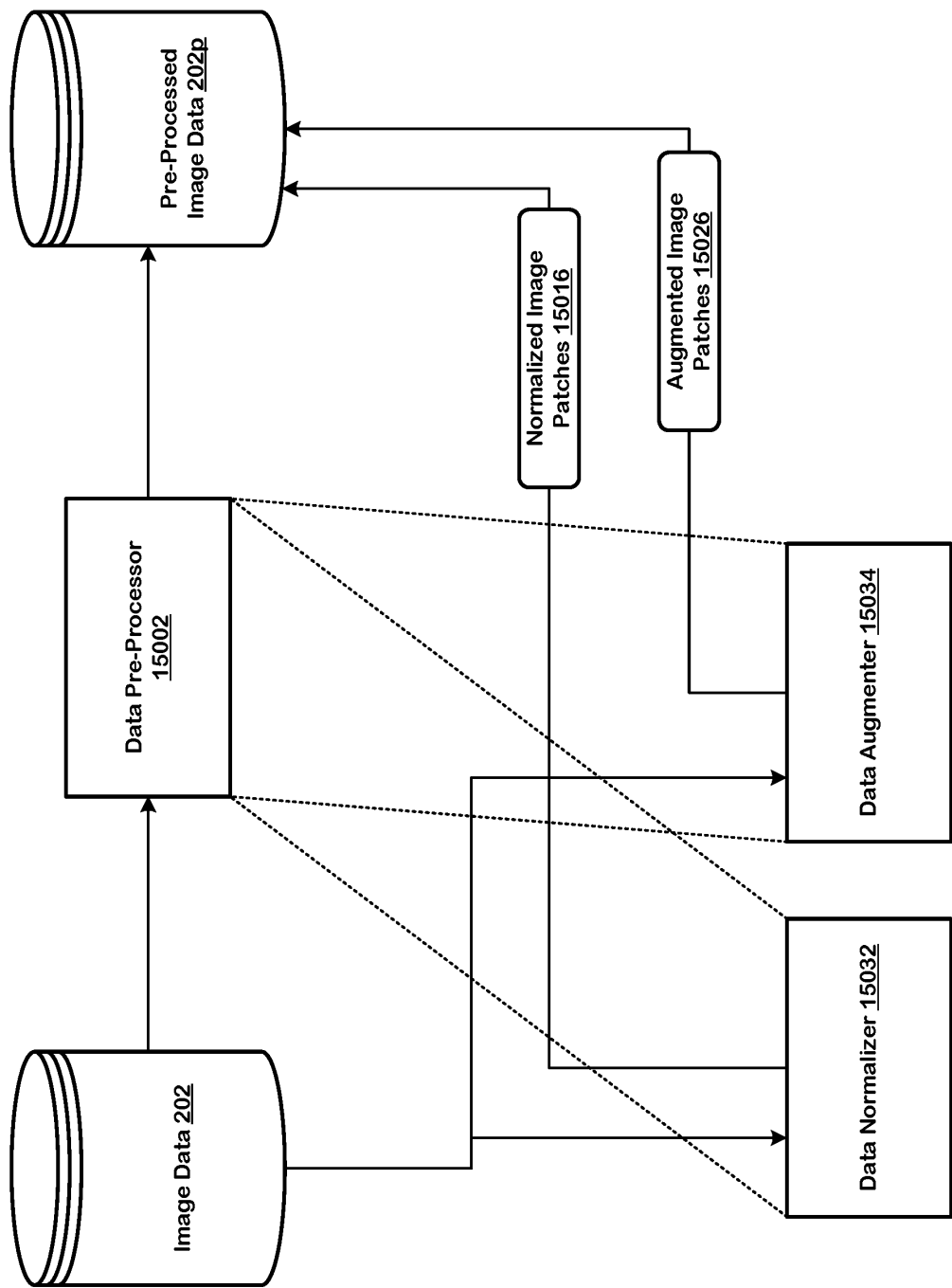


Figure 150

Train on Bacteria, Test on Human (same assay)

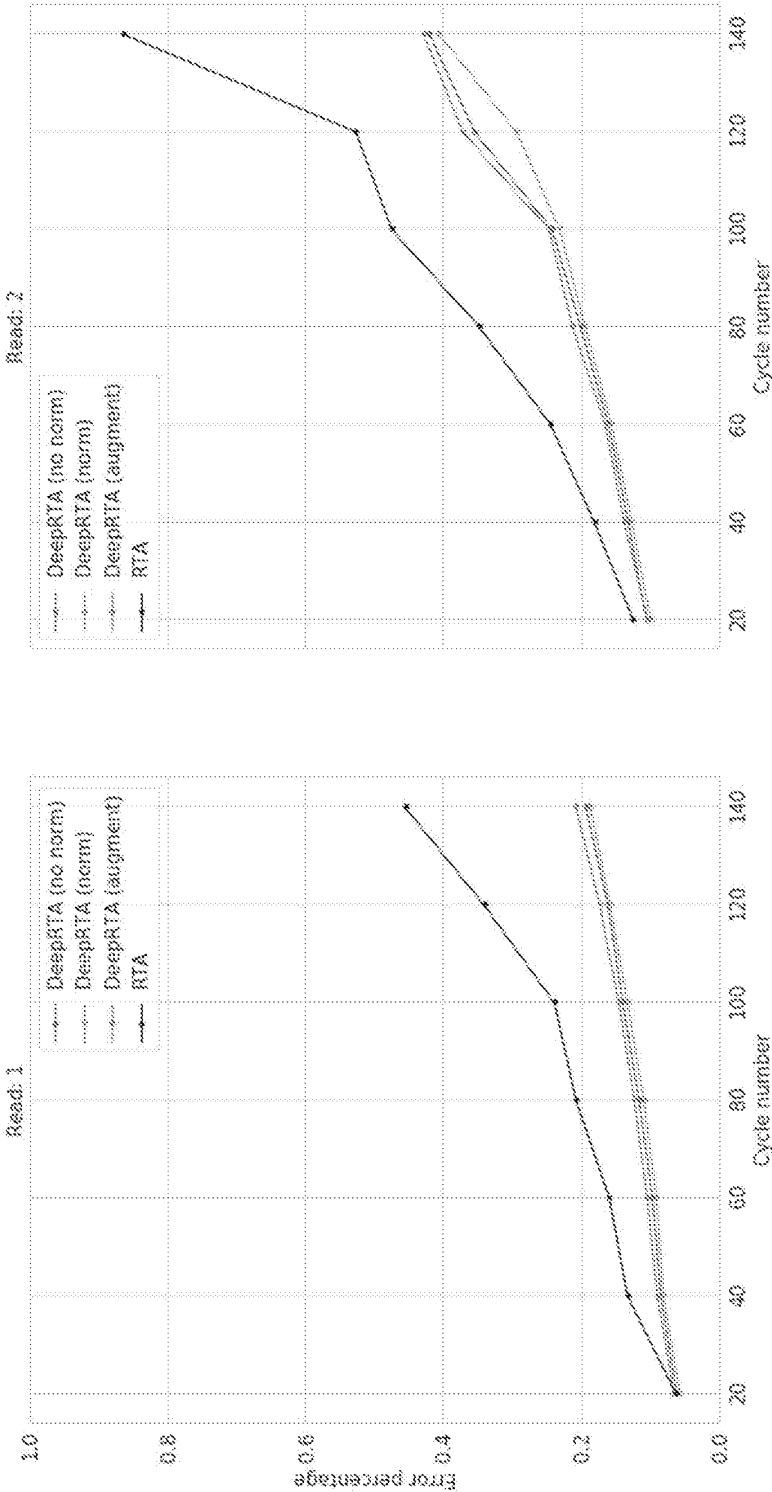


Figure 151

Assay generalization – Exome sequencing

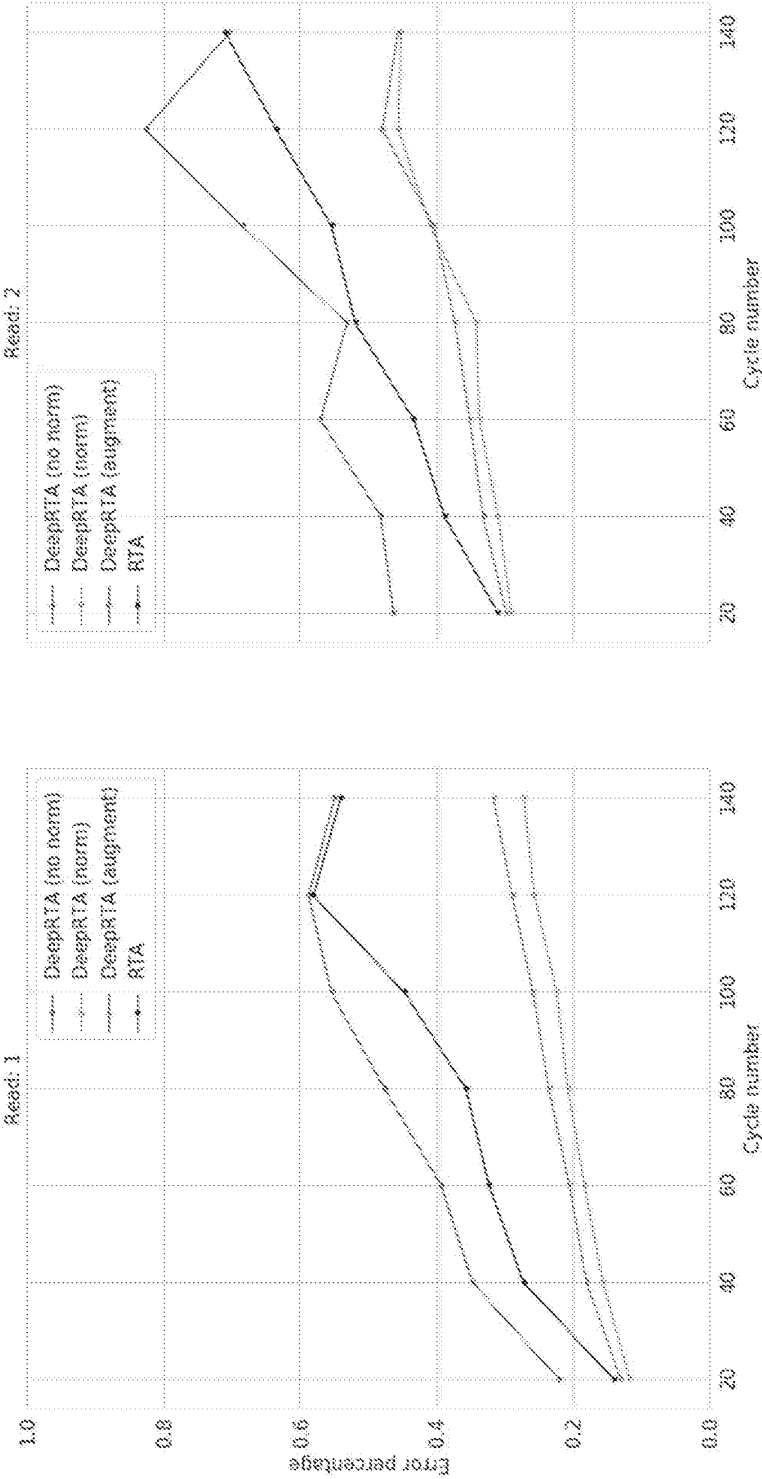


Figure 152