CLAIMS

[Claim 1]. A computerimplemented method of determining cluster metadata, including cluster background, cluster centers, and cluster shapes, from image data generated based upon one or more clusters, the method including:

receiving as input, the image data derived from a sequence of images,

wherein each image in the sequence of images represents an imaged region and depicts intensity emissions of the one or more clusters and their surrounding background at a respective one of a plurality of sequencing cycles of a sequencing run, and

wherein the image data comprises image patches extracted from each image in the sequence of images;

processing the image data through a neural network to generate a feature map of the image data, wherein the neural network is trained for cluster metadata determination task, including determining cluster background, cluster centers, and cluster shapes;

processing the feature map through an output layer to generate an output indicating properties of respective portions of the imaged region;

thresholding classification scores of the output values and classifying a first subset of the respective portions of the imaged region as background portions depicting the surrounding background;

locating peaks in the classification scores and classifying a second subset of the respective portions of the imaged region as containing centers of the clusters; and

applying a segmenter to classification scores and determining shapes of the clusters as non-overlapping and contiguous portions of the imaged region.

[Claim 2]. The computer-implemented method of claim 1, wherein the properties include

whether a portion represents background or cluster, and

whether a portion represents a center of a plurality of contiguous image portions each representing a same cluster.

[Claim 3]. The computer-implemented method of claim 1, wherein the output identifies

the one or more clusters, whose intensity emissions are depicted by the input image data, as disjoint regions of adjoining units,

centers of the one or more clusters as center units at centers of mass of the respective ones of the disjoint regions, and their surrounding background as background units not belonging to any of the disjoint regions.

[Claim 4]. The computer-implemented method of claim 3, wherein the adjoining units in the respective ones of the disjoint regions have intensity values weighted according to distance of an adjoining unit from a center unit in a disjoint region to which the adjoining unit belongs.

[Claim 5]. Thecomputer-implemented method of any of claims 1-4, wherein the output is a binary map which classifies each portion as cluster or background.

[Claim 6]. The computer-implemented method of any of claims 1-5, wherein the output is a ternary map which classifies each portion as cluster, background, or center.

[Claim 7]. The computer-implemented method of any of claims 1-6, further including:

applying a peak locator to the output to find peak intensities in the output;

determining location coordinates of the centers of the clusters based on the peak intensities;

downscaling the location coordinates by an upsampling factor used to prepare the input image data; and

storing the downscaled location coordinates in memory for use in base calling the clusters.

[Claim 8]. Thecomputer-implemented method of any of claims 1-7, further including:

categorizing the adjoining units in the respective ones of the disjoint regions as cluster interior units belonging to a

same cluster; and

storing the categorization and downscaled location coordinates of the cluster interior units in the memory on a cluster-by-cluster basis for use in base calling the clusters.

[Claim 9]. Thecomputer-implemented method of any of claims 1-8, further including:

obtaining training data for training the neural network,

wherein the training data includes a plurality of training examples and corresponding ground truth data,

wherein each training example includes image data from a sequence of image sets,

wherein each image in the sequence of image sets represents a tile of a flow cell and depicts intensity emissions of clusters on the tile and their surrounding background captured for a particular image channel at a particular one of a plurality of sequencing cycles of a sequencing run performed on the flow cell, and

wherein each ground truth data identifies properties of respective portions of the training examples; and

using a gradient descent training technique to train the neural network and generating outputs for the training examples that progressively match the ground truth data, including iteratively

optimizing a loss function that minimizes error between the outputs and the ground truth data, and

updating parameters of the neural network based on the error.

[Claim 10]. The computer-implemented method of any of claims 1-9, wherein the properties comprise

clusters, whose intensity emissions are depicted by the image data of a corresponding training example, as disjoint regions of adjoining units,

centers of the clusters as center units at centers of mass of the respective ones of the disjoint regions, and

their surrounding background as background units not belonging to any of the disjoint regions.

[Claim 11]. The computer-implemented method of any of claims 1-10, wherein the properties comprise identifying whether a unit is a center or a non-center.

[Claim 12]. The computer-implemented method of claim 9, further including:

upon error convergence after a final iteration, storing the updated parameters of the neural network in memory to be applied to further neural network-based template generation and base calling.

[Claim 13]. The computer-implemented method of any of claims 9 or 12, wherein, in the ground truth data, the adjoining units in the respective ones of the disjoint regions have intensity values weighted according to distance of an adjoining unit from a center unit in a disjoint region to which the adjoining unit belongs.

[Claim 14]. The computer-implemented method of any of claims 9 or 12-13, wherein, in the ground truth data, the center units have highest intensity values within the respective ones of the disjoint regions.

[Claims 15]. The computer-implemented method of any of claims 9or 12-14, wherein the loss function is mean squared error and the error is minimized on a unit-basis between the normalized intensity values of corresponding units in the outputs and the ground truth data.

[Claim 16]. The computer-implemented method of any of claims 9or 12-15, wherein, in the training data, multiple training examples respectively include as image data different image patches from each image in a sequence of image sets of a same tile, and

wherein at least some of the different image patches overlap with each other.

[Claim 17]. The computer-implemented method of any of claims 9 or 12-16, wherein, in the ground truth data,

units classified as cluster centers are all assigned a same first predetermined class score, and

units classified as non-centers are all assigned a same second predetermined class score.

[Claim 18]. The computer-implemented method of any of claims 9 or 12-17, wherein the loss function is custom weighted binary cross entropy loss and the error is minimized on a unit-basis between predicted classification scores and the class scores of corresponding units in the outputs and the ground truth data.

[Claim 19]. The computer-implemented method of any of claims 9 or 12-18, wherein, in the ground truth data,

units classified as background are all assigned a same first predetermined class score,

units classified as cluster centers are all assigned a same second predetermined class score, and

units classified as cluster interior are all assigned a same third predetermined class score.

[Claim 20]. The computer-implemented method of any of claims 1-19, further including:

thresholding output values of the units and classifying a first subset of the units as background units depicting the surrounding background;

locating peaks in the output values of the units and classifying a second subset of the units as center units containing centers of the clusters; and

applying a segmenter to the output values of the units and determining shapes of the clusters as non-overlapping regions of contiguous units separated by the background units and centered at the center units, wherein the segmenter begins with the center units and determines, for each center unit, a group of successively contiguous units that depict a same cluster whose center is contained in the center unit.

[Claim 21]. The computer-implemented method of any of claims 1-20, wherein the non-overlapping regions have irregular contours, further including:

determining cluster intensity of a given cluster by:

identifying units that contribute to the cluster intensity of the given cluster based on a corresponding non-overlapping region of contiguous units that identifies a shape of the given cluster;

locating the identified units in one or more optical, pixel resolution images generated for one or more image channels at a current sequencing cycle;

in each of the images, interpolating intensities of the identified units, combining the interpolated intensities, and normalizing the combined interpolated intensities to produce a per-image cluster intensity for the given cluster in each of the images; and

combining the per-image cluster intensity for each of the images to determine the cluster intensity of the given cluster at the current sequencing cycle.

[Claim 22]. The computer-implemented method of any of claims 1-21, wherein the non-overlapping regions have irregular contours, further including:

determining cluster intensity of a given cluster by:

identifying units that contribute to the cluster intensity of the given cluster based on a corresponding non-overlapping region of contiguous units that identifies a shape of the given cluster;

locating the identified units in one or more unit resolution images upsampled from corresponding optical, pixel resolution images generated for one or more image channels at a current sequencing cycle;

in each of the upsampled images, combining intensities of the identified units and normalizing the combined intensities to produce a per-image cluster intensity for the given cluster in each of the upsampled images; and

combining the per-image cluster intensity for each of the upsampled images to determine the cluster intensity of the given cluster at the current sequencing cycle.

[Claim 23]. The computer-implemented method of any of claims 1-22, wherein the normalizing is based on a normalization factor, and

wherein the normalization factor is a number of the identified units.

[Claim 24]. The computer-implemented method of any of claims 1-23, further including:

base calling the given cluster based on the cluster intensity at the current sequencing cycle.

[Claim 25]. A computer-implemented method of determining metadata about clusters on a flow cell, the method including:

accessing image data that depicts intensity emissions of the clusters;

processing the image data through one or more layers of a neural network and generating a feature map of the image data; and

processing the feature map through an output layer and generating an output that identifies shapes and sizes of the clusters and centers of the clusters.

[Claim 26]. The computer-implemented method of claim 25, wherein the image data further depicts intensity emissions of surrounding background of the clusters, further including:

the output identifying spatial distribution of the clusters on the flow cell, including the surrounding background and boundaries between the clusters.

[Claim 27]. A computer-implemented method, including:

processing image data through a neural network and generating an alternative representation of the image data, wherein the image data depicts intensity emissions of clusters; and

processing the alternative representation through an output layer and generating an output that identifies metadata about the clusters, including at least one of spatial distribution of the clusters, shapes of the clusters, centers of the clusters, and/or boundaries between the clusters.