Appendix 1

This appendix includes the details of FLocK features, where the pre-defined cluster number is set to 2. Features belonging to Sets A and B are included below. For more details, we direct the reader to the code we released on Github https://github.com/hacylu/FLocK

A. Features that do not consider cluster types (i.e. Set A)

Feature set A includes four groups, A1 to A4, which are described below:

Group A1 - Intersection between different FLocKs: intersection properties/measurement of FLocKs or overlapped area portion between proximal FLocKs.

Feature name	Description
FLocK - portion of intersected FLocK	(total number of intersected FLocKs) / (total number of FLocKs)
FLocK-abs. number of intersected FLocK	absolute number of intersected FLocKs
FLock-portion of highly intersected FLock:T=0.1 FLock-portion of highly intersected FLock:T=0.2 FLock-portion of highly intersected FLock:T=0.3 FLock-portion of highly intersected FLock:T=0.4 FLock-portion of highly intersected FLock:T=0.5 FLock-portion of highly intersected FLock:T=0.6 FLock-portion of highly intersected FLock:T=0.7 FLock-portion of highly intersected FLock:T=0.8 FLock-portion of highly intersected FLock:T=0.9	Denote two FLocKs as C and D , corresponding areas are denoted as $area(C)$ and $area(D)$, respectively. If two FLocKs intersect, the corresponding area is denoted as $area(C \cap D)$. The ratio of the intersected region is calculated as $P = area(C \cap D)/area(C)$ or $P = area(C \cap D)/area(D)$. We calculate the portion of intersected FLocKs whose $P > T$, where T is a set of pre-defined thresholds, ranging from 0.1 to 0.9 in increments of 0.1. For example, in an image, we have 30 FLocKs and 15 of them have $P > 0.2$. The feature 'FLocK-portion of highly intersected FLocK: $T = 0.2$ ' value is $15/30 = 0.5$.
FLocK-number of highly intersected FLocK:T=0.1 FLocK-number of highly intersected FLocK:T=0.2 FLocK-number of highly intersected FLocK:T=0.3 FLocK-number of highly intersected FLocK:T=0.4 FLocK-number of highly intersected FLocK:T=0.5 FLocK-number of highly intersected FLocK:T=0.6 FLocK-number of highly intersected FLocK:T=0.7 FLocK-number of highly intersected FLocK:T=0.8 FLocK-number of highly intersected FLocK:T=0.9	Denote two FLocks as C and D , corresponding areas are denoted as $area(C)$ and $area(D)$, respectively. If two FLocks intersect, the corresponding area is denoted as $area(C \cap D)$. The ratio of the intersected region is calculated as $P = area(C \cap D)/area(C)$ or $P = area(C \cap D)/area(D)$. We calculate the portion of intersected FLocks whose $P > T$, where T is a set of pre-defined thresholds, ranging from 0.1 to 0.9 in increments of 0.1. For example, in an image, we have 30 FLocks and 15 of them have $P > 0.2$. 'FLock-number of highly intersected FLock: $T = 0.2$ ' value is 15.
FLocK-mean(intersection area portion:small) FLocK-median(intersection area portion:small) FLocK-std(intersection area portion:small) FLocK-range(intersection area portion:small) FLocK-min(intersection area portion:small) FLocK-max(intersection area portion:small)	This set of features measures statistics of intersecting FLocKs. Denote two FLocKs as C and D, corresponding areas are denoted as area(C) and area(D), respectively. The intersection is designated as C ∩ D. Feature (intersection area portion:large) is calculated as

FLocK-kurtosis(intersection area portion:small) FLocK-skewness(intersection area portion:small) FLocK-mean(intersection area portion:large) FLocK-median(intersection area portion:large) FLocK-std(intersection area portion:large) FLocK-range(intersection area portion:large) FLocK-min(intersection area portion:large) FLocK-max(intersection area portion:large) FLocK-kurtosis(intersection area portion:large) FLocK-skewness(intersection portion:large) FLocK-mean(intersection area portion:mean) FLocK-median(intersection area portion:mean) FLocK-std(intersection area portion:mean) FLocK-range(intersection area portion:mean) FLocK-min(intersection area portion:mean) FLocK-max(intersection area portion:mean) FLocK-kurtosis(intersection area portion:mean) FLocK-skewness(intersection area area(C&D)/max(area(C), area(D)); (intersection area portion:small) is calculated area(C&D)/min(area(C), area(D)); (intersection portion:mean) calculated area is area(C&D)/mean(area(C), area(D)). Since we have multiple FLocKs in an image, we use 8 types of statistics (mean, median, standard deviation, range, minimum value, maximum value, kurtosis, and skewness), to characterize the whole image (this applies to the following features as well).

FLocK-mean(intersection area abs val)

FLocK-median(intersection area abs val)

portion:mean)

FLocK-std(intersection area abs val)
FLocK-range(intersection area abs val)
FLocK-min(intersection area abs val)
FLocK-max(intersection area abs val)
FLocK-kurtosis(intersection area abs val)
FLocK-skewness(intersection area abs val)

Denote two FLocks as C and D, areas of these two FLocks are calculated as area(C) and area(D), respectively. The intersected region of C and D is calculated as $C \cap D$. Feature intersection area abs val is calculated as area(C&D). Since we have multiple FLocks in an image, we use 8 common statistics (mean, median, standard deviation, range, minimum value, maximum value, kurtosis, and skewness), to quantify the whole image (this applies to the following features as well).

Group A2 – Size of FLocKs: statistics of the size of FLocK/number of nuclei within a FLocK (include the single cell as a FLocK) across the whole image.

Feature name	Description
FLocK-mean(size of all FLocK)	These features capture the size of FLocKs. Size
FLocK-median(size of all FLocK)	of all FLocK is self explanatory. Size of non-1-
FLocK-std(size of all FLocK)	2cell FLocK, represents the size of all FLocKs
FLocK-range(size of all FLocK)	containing more than 2 cells.
FLocK-min(size of all FLocK)	
FLocK-max(size of all FLocK)	
FLocK-kurtosis(size of all FLocK)	
FLocK-skewness(size of all FLocK)	
FLocK-mean(size of non-1-2cell FLocK)	
FLocK-median(size of non-1-2cell FLocK)	
FLocK-std(size of non-1-2cell FLocK)	
FLocK-range(size of non-1-2cell FLocK)	
FLocK-min(size of non-1-2cell FLocK)	
FLocK-max(size of non-1-2cell FLocK)	

FLocK-kurtosis(size of non-1-2cell FLocK) FLocK-skewness(size of non-1-2cell FLocK)	
FLocK-non-1-2cell FLocK ratio	(number of FLocKs containing more than two cells) / (number of FLocKs in the image)
FLocK-mean(nuclei no. in FLocK)	Number of nuclei in each FLocK.
FLocK-median(nuclei no. in FLocK)	
FLocK-std(nuclei no. in FLocK)	
FLocK-range(nuclei no. in FLocK)	
FLocK-min(nuclei no. in FLocK)	
FLocK-max(nuclei no. in FLocK)	
FLocK-kurtosis(nuclei no. in FLocK)	
FLocK-skewness(nuclei no. in FLocK)	

Group A3 – Disorder of nuclear morphology: variation in distance/size/color/shape of nucleus compared to the centroid of a FLocK (FLocK has >=2 cells).

Feature name	Description
'FLocK-mean(var of nuclear feat1to the centroid)'	This set of features measures the variation in nuclear features all nodes within the flock with
'FLocK-median(var of nuclear feat1to the centroid)'	respect to the node corresponding to the centroid of the FLock. If the FLocks were
'FLocK-std(var of nuclear feat1to the centroid)'	constructed in a 2-nuclear feature space, feat1
'FLocK-range(var of nuclear feat1to the centroid)'	and <i>feat2</i> would represent the first and second nuclear features, respectively. Denote the
'FLocK-min(var of nuclear feat1to the centroid)'	centroid of FLocK in feature space 1 or 2 as F,
'FLocK-max(var of nuclear feat1to the centroid)'	each nuclear feature as N _i , the variation is
'FLocK-kurtosis(var of nuclear feat1to the centroid)'	calculated as $\sqrt{\frac{\sum_{i}^{Q}(N_{i}-F)^{2}}{o}}$, where Q is the
'FLocK-skewness(var of nuclear feat1to the centroid)'	number of nuclei in the FLocK.
'FLocK-mean(var of nuclear feat2to the centroid)'	
'FLocK-median(var of nuclear feat2to the centroid)'	
'FLocK-std(var of nuclear feat2to the centroid)'	
'FLocK-range(var of nuclear feat2to the centroid)'	
'FLocK-min(var of nuclear feat2to the centroid)'	
'FLocK-max(var of nuclear feat2to the centroid)'	
'FLocK-kurtosis(var of nuclear feat2to the centroid)'	
'FLocK-skewness(var of nuclear feat2to the centroid)'	
'FLocK-mean(var of cell to FLocK centroid)'	This set of features measures the variation
'FLocK-median(var of cell to FLocK centroid)'	between the nuclear physical distance and the
'FLocK-std(var of cell to FLocK centroid)'	centroid of the FLocK. Denote the centroid of
'FLocK-range(var of cell to FLocK centroid)'	FLocK as M, each nuclear centroid as N _i , the
'FLocK-min(var of cell to FLocK centroid)'	variation is calculated as $\sqrt{\frac{\sum_{l}^{Q}(N_{l}-M)^{2}}{o}}$, where Q
'FLocK-max(var of cell to FLocK centroid)'	variation is calculated as $\sqrt{\frac{-i}{Q}}$, where Q
'FLocK-kurtosis(var of cell to FLocK centroid)' 'FLocK-skewness(var of cell to FLocK centroid)'	is the number of nuclei in the FLocK.

Group A4 – Spatial arrangement of FLocKs: use the centroids of FLocK as nodes, we build global graph (Voronoi diagram, Delaunay Triangulation, and Minimum Spanning Trees) to connect centroids of different FLocKs. We have another set of features that use the centroids of intersected regions of FLocKs as nodes which we have not listed here.

Feature name	Description
FLocK-FLocK centroid-Area Standard Deviation	Use the centroids of FLocKs as nodes to
FLocK-FLocK centroid-Area Average	construct Voronoi Diagram. Voronoi Diagram
FLocK-FLocK centroid-Area Minimum /	based features related to the area, perimeter,
Maximum	and chord of Voronoi cells are calculated.
FLocK-FLocK centroid-Area Disorder	
FLocK-FLocK centroid-Perimeter Standard	
Deviation	
FLocK-FLocK centroid-Perimeter Average	
FLocK-FLocK centroid-Perimeter Minimum /	
Maximum	
FLocK-FLocK centroid-Perimeter Disorder	
FLocK-FLocK centroid-Chord Standard	
Deviation	
FLocK-FLocK centroid-Chord Average	
FLocK-FLocK centroid-Chord Minimum /	
Maximum	
FLocK-FLocK centroid-Chord Disorder	
FLocK-FLocK centroid-Side Length Minimum /	Use the centroids of FLocKs as nodes to
Maximum	construct Delaunay Triangulation. Delaunay
FLocK-FLocK centroid-Side Length Standard	Triangulation based-features related to the
Deviation	triangle edge length, and triangle area are
FLocK-FLocK centroid-Side Length Average	calculated.
FLocK-FLocK centroid-Side Length Disorder	
FLocK-FLocK centroid-Triangle Area Minimum /	
Maximum	
FLocK-FLocK centroid-Triangle Area Standard	
Deviation	
FLocK-FLocK centroid-Triangle Area Average	
FLocK-FLocK centroid-Triangle Area Disorder	
FLocK-FLocK centroid-MST Edge Length	Use the centroids of FLocKs as nodes to
Average	construct Minimum Spanning Tree (MST). MST
FLocK-FLocK centroid-MST Edge Length	based-features related to the edge length are
Standard Deviation	calculated.
FLocK-FLocK centroid-MST Edge Length	
Minimum / Maximum	
FLocK-FLocK centroid-MST Edge Length	
Disorder	
FLocK-FLocK centroid-Average distance to 3	Calculate Euclidean distances between
Nearest Neighbors	centroids of FLocKs and centroids of other
FLocK-FLocK centroid-Average distance to 5	FLocKs. Euclidean distance based-features,
Nearest Neighbors	including the statistics of K-nearest neighbors,
FLocK-FLocK centroid-Average distance to 7	and nearest neighbors in a certain pixel radius
Nearest Neighbors	are then calculated.
FLocK-FLocK centroid-Standard Deviation	
distance to 3 Nearest Neighbors	

FLocK-FLocK centroid-Standard **Deviation** distance to 5 Nearest Neighbors FLocK-FLocK centroid-Standard **Deviation** distance to 7 Nearest Neighbors FLocK-FLocK centroid-Disorder of distance to 3 **Nearest Neighbors** FLocK-FLocK centroid-Disorder of distance to 5 **Nearest Neighbors** FLocK-FLocK centroid-Disorder of distance to 7 **Nearest Neighbors** FLocK-FLocK centroid-Avg. Nearest Neighbors in a 10 Pixel Radius FLocK-FLocK centroid-Avg. Nearest Neighbors in a 20 Pixel Radius FLocK-FLocK centroid-Avg. Nearest Neighbors in a 30 Pixel Radius FLocK-FLocK centroid-Avg. Nearest Neighbors in a 40 Pixel Radius FLocK-FLocK centroid-Avg. Nearest Neighbors in a 50 Pixel Radius FLocK-FLocK centroid-Standard **Deviation Nearest Neighbors in a 10 Pixel Radius** centroid-Standard **Deviation** FLocK-FLocK Nearest Neighbors in a 20 Pixel Radius FLocK-FLocK centroid-Standard Deviation Nearest Neighbors in a 30 Pixel Radius **Deviation** FLocK-FLocK centroid-Standard Nearest Neighbors in a 40 Pixel Radius FLocK-FLocK centroid-Standard **Deviation** Nearest Neighbors in a 50 Pixel Radius FLocK-FLocK centroid-Disorder of Nearest Neighbors in a 10 Pixel Radius FLocK-FLocK centroid-Disorder of Nearest Neighbors in a 20 Pixel Radius FLocK-FLocK centroid-Disorder of Nearest Neighbors in a 30 Pixel Radius FLocK-FLocK centroid-Disorder of Nearest Neighbors in a 40 Pixel Radius FLocK-FLocK centroid-Disorder of Nearest Neighbors in a 50 Pixel Radius

FLocK-FLocK centroid-Area of polygons
FLocK-FLocK centroid-Number of nuclei
FLocK-FLocK centroid-Density of Nuclei
FLocK-mean(density of non-1-2cell FLocKs)
FLocK-median(density of non-1-2cell FLocKs)
FLocK-std(density of non-1-2cell FLocKs)
FLocK-range(density of non-1-2cell FLocKs)
FLocK-min(density of non-1-2cell FLocKs)
FLocK-max(density of non-1-2cell FLocKs)
FLocK-kurtosis(density of non-1-2cell FLocKs)
FLocK-skewness(density of non-1-2cell FLocKs)

Calculate the density of a FLock. The density is calculated as: (the number of nuclei in FLock) / (area of the FLock). Statistics of all FLocks in an image are calculated as well.

FLocK-mean(density of all FLocKs)

FLocK-median(density of all FLocKs)

FLocK-std(density of all FLocKs)

FLocK-range(density of all FLocKs)

FLocK-min(density of all FLocKs)

FLocK-max(density of all FLocKs)

FLocK-kurtosis(density of all FLocKs)

FLocK-skewness(density of all FLocKs)

B \ Features that consider cluster types (for example, cluster type number=2, we denote the first type as type 1, the second type as type 2)

Group B1 - Intersection between different types of FLocKs: intersection properties/measurement of FLocKs corresponding to different cluster types.

Feature name	Description
FLocK-portion of intra-type FLocK intersction-	These features measure the interaction
Type1	between different types of FLocKs. 'Intra-type'
FLocK-portion of inter-type FLocK intersction-	indicates the FLocKs that belonged to the same
Type1	type (i.e. either Type1 or Type2), whereas
FLocK-portion of intra-type FLocK intersction-	'Inter-type' refers to the FLocKs between
Type2	different types (e.g. between Type 1 and Type
FLocK-portion of inter-type FLocK intersction-	2). Denote the total number of FLocKs in an
Type2	image as N, the number of FLocKs belonging to
FLocK-number of intra-type FLocK intersction-	type 1 as N ₁ , the number of FLocKs belonging to
Type1	type 2 as N ₂ , the number of Type 1 FLocKs that
FLocK-number of inter-type FLocK intersction-	intersect with Type 1 Flocks as N_1^1 , the number
Type1	of Type 1 FLocKs that intersect with Type 2
FLocK-number of intra-type FLocK intersction-	Flocks as N_1^2 . Feature portion of intra-type
Type2	FLocK intersction-Type1, is calculated as N_1^1
FLocK-number of inter-type FLocK intersction-	N_1 .
Type2	Feature portion of inter-type FLocK intersction-
	<i>Type1</i> , is calculated as N_1^2/N_1 .

Group B2 - Enrichment of the K-nearest FLocKs: quantify the diversity of FLocKs, with pre-defined phenotype number, in terms of nearest FLocKs.

FLocK-mean(portion of other FLocK ty	pes The po	ortion of other types of FLocKs within 5,
in5nearest neighbors)	10, 15	nearest neighbored FLocKs. For example,
FLocK-median(portion of other FLocK ty	es within	10 nearest neighbored FLocKs, 3 of them
in5nearest neighbors)	are fro	om the same type, and the other 7 are
FLocK-std(portion of other FLocK ty	es from t	he other type. The portion of the other
in5nearest neighbors)	types	of FLocKs within 10 nearest neighbored
FLocK-range(portion of other FLocK ty	es FLocKs	is 0.7.
in5nearest neighbors)		
FLocK-min(portion of other FLocK ty	oes	
in5nearest neighbors)		
FLocK-max(portion of other FLocK ty	oes	
in5nearest neighbors)		
FLocK-kurtosis(portion of other FLocK ty	oes	
in5nearest neighbors)		

FLocK-skewness(portion of other FLocK types in5nearest neighbors) FLocK-mean(portion of other FLocK types in10nearest neighbors) FLocK-median(portion of other FLocK types *in10nearest neighbors)* FLocK-std(portion of other FLocK types in10nearest neighbors) FLocK-range(portion of other FLocK types *in10nearest neighbors)* FLocK-min(portion of other FLocK types in10nearest neighbors) FLocK-max(portion of other FLocK types in10nearest neighbors) FLocK-kurtosis(portion of other FLocK types in10nearest neighbors) FLocK-skewness(portion of other FLocK types in10nearest neighbors) FLocK-mean(portion of other FLocK types *in15nearest neighbors*) FLocK-median(portion of other FLocK types *in15nearest neighbors*) FLocK-std(portion of other FLocK types in15nearest neighbors) FLocK-range(portion of other FLocK types *in15nearest neighbors*) FLocK-min(portion of other FLocK types *in15nearest neighbors*) FLocK-max(portion of other FLocK types *in15nearest neighbors*) FLocK-kurtosis(portion of other FLocK types *in15nearest neighbors)* FLocK-skewness(portion of other FLocK types in15nearest neighbors)

Group B3 – The spatial arrangement of a specific type of FLocKs: use the centroids of a specific type of FLocK (e.g. Type1 or Type2) as nodes, we build global graph (Voronoi diagram, Delaunay Triangulation, and minimum spanning trees) to extract global graph features. Here we consider Type 1 FLocKs only.

Feature name	Description
FLocK-Type1FLocK centroid-Area Standard	Use centroid of a Type 1 FLocK to construct
Deviation	Voronoi Diagram by connecting the centroids of
FLocK-Type1FLocK centroid-Area Average	all Type 1 FLocKs. Voronoi Diagram based-
FLocK-Type1FLocK centroid-Area Minimum /	features related to the area, perimeter, and
Maximum	chord of Voronoi cells are calculated.
FLocK-Type1FLocK centroid-Area Disorder	
FLocK-Type1FLocK centroid-Perimeter Standard	
Deviation	
FLocK-Type1FLocK centroid-Perimeter Average	

FLocK-Type1FLocK centroid-Perimeter	
Minimum / Maximum	
FLocK-Type1FLocK centroid-Perimeter Disorder	
FLocK-Type1FLocK centroid-Chord Standard	
Deviation	
FLocK-Type1FLocK centroid-Chord Average	
FLocK-Type1FLocK centroid-Chord Minimum /	
Maximum	
FLocK-Type1FLocK centroid-Chord Disorder	
	Use Type 1 FLocK centroid to construct
	Delaunay Triangulation by connecting the
FLocK-Type1FLocK centroid-Side Length c	centroids of all Type 1 FLocKs. Delaunay
Standard Deviation	Triangulation based-features related to the
FLocK-Type1FLocK centroid-Side Length t	triangle side length, and triangle area are
	calculated.
FLocK-Type1FLocK centroid-Side Length	
Disorder	
FLocK-Type1FLocK centroid-Triangle Area	
Minimum / Maximum	
FLocK-Type1FLocK centroid-Triangle Area	
Standard Deviation	
FLocK-Type1FLocK centroid-Triangle Area	
Average	
FLocK-Type1FLocK centroid-Triangle Area	
Disorder	
FLocK-Type1FLocK centroid-MST Edge Length \	Use Type 1 FLocK centroid to construct
Average	Minimum Spanning Tree (MST) by connecting
FLocK-Type1FLocK centroid-MST Edge Length t	the centroids of all Type 1 FLocKs. MST based-
Standard Deviation f	features related to the edge length are
FLocK-Type1FLocK centroid-MST Edge Length c	calculated.
Minimum / Maximum	
FLocK-Type1FLocK centroid-MST Edge Length	
Disorder	
,, ,,	Calculate Euclidean distances between
FLocK-Type1FLocK centroid-Number of nuclei	centroids of Type 1 FLocKs and centroids of
, , ,	other FLocKs. Euclidean distance based-
1	features, including statistics of K-nearest
3	neighbors, and nearest neighbors in a
FLocK-Type1FLocK centroid-Average distance to	certain pixel radius.
5 Nearest Neighbors	
FLocK-Type1FLocK centroid-Average distance to	
7 Nearest Neighbors	
FLocK-Type1FLocK centroid-Standard Deviation	
distance to 3 Nearest Neighbors	
FLocK-Type1FLocK centroid-Standard Deviation	
distance to 5 Nearest Neighbors	
FLocK-Type1FLocK centroid-Standard Deviation	
Make a section 7 Africa and Africa 4.4	
distance to 7 Nearest Neighbors	l l
fistance to 7 Nearest Neighbors FLocK-Type1FLocK centroid-Disorder of distance to 3 Nearest Neighbors	

FLocK-Type1FLocK centroid-Disorder of distance to 5 Nearest Neighbors centroid-Disorder of FLocK-Type1FLocK distance to 7 Nearest Neighbors FLocK-Type1FLocK centroid-Avg. Nearest Neighbors in a 10 Pixel Radius FLocK-Type1FLocK centroid-Avg. Nearest Neighbors in a 20 Pixel Radius FLocK-Type1FLocK centroid-Avg. Nearest Neighbors in a 30 Pixel Radius FLocK-Type1FLocK centroid-Avg. Nearest Neighbors in a 40 Pixel Radius FLocK-Type1FLocK centroid-Avg. Nearest Neighbors in a 50 Pixel Radius FLocK-Type1FLocK centroid-Standard Deviation Nearest Neighbors in a 10 Pixel Radius FLocK-Type1FLocK centroid-Standard Deviation Nearest Neighbors in a 20 Pixel Radius FLocK-Type1FLocK centroid-Standard Deviation Nearest Neighbors in a 30 Pixel Radius FLocK-Type1FLocK centroid-Standard Deviation Nearest Neighbors in a 40 Pixel Radius FLocK-Type1FLocK centroid-Standard Deviation Nearest Neighbors in a 50 Pixel Radius FLocK-Type1FLocK centroid-Disorder of Nearest Neighbors in a 10 Pixel Radius FLocK-Type1FLocK centroid-Disorder of Nearest Neighbors in a 20 Pixel Radius FLocK-Type1FLocK centroid-Disorder of Nearest Neighbors in a 30 Pixel Radius FLocK-Type1FLocK centroid-Disorder of Nearest Neighbors in a 40 Pixel Radius FLocK-Type1FLocK centroid-Disorder of Nearest Neighbors in a 50 Pixel Radius