Settings to generate a generic synthetic dataset

Antonio Martinez-Sanchez[1,*]

May 18, 2023

[1]Department of Information and Communications Engineering, Faculty of Computers Sciences, University of Murcia, Campus de Espinardo 30100 Murcia, Spain

ACRONYM	MEANING
cryo-ET	cryo-Electron Tomography
3D	Three dimensions
DL	Deep Learning
CNN	Convolutional Neuronal Networks
TM	Template Matching
2D	Two dimensions
TEM	Transmission Electron Microscopy
VOI	Volume Of Interest
SAWLC	Self-Avoiding Worm-Like Chain
PDB	Protein Data Bank
CSR	Complete Spatial Randomness
SNR	Signal-to-Noise Ratio
MRC	Medical Research Council
VTK	Visualization ToolKit
CSV	Comma Separated Values
MT	MicroTubule

Table 1: Glossary of acronyms by order of appearance in the main publication.

SCOPE	SYMBOL	LABEL
Membrane	$\mathcal{D}$	Cellular context scalar field (or density map)
	$\mathcal{V}$	VOI
	t	Membrane thickness or distance between both layers
	σ_l	Membrane layer thickness
	$\hat{L}$	Non-smooth membrane layer set
	L	Smoothed membrane scalar field
	c	Membrane structural model center
	r	Spherical membrane radius
	a	Ellipsoid membrane semi-axis lengths
	ϵ	Maximum ellipsoid eccentricity
	d	Toroid membrane major radius
	e	Toroid membrane minor radius
	$\mathcal{G}$	Gaussian filter for 3D scalar fields
	λ	Hessian matrix major eigenvalue
	$\mathcal{L}$	Membrane surface
Filament	M	Structural unit 3D scalar field
	τ	Curve (filament center-line) tangent vector
	d_M	Distance between two structural units in a polymer
	s	Curve length
	p	Polymer persistence length
	κ	Curve curvature
	f	Helical curve circular component
	h_z	Helical curve elevation in Z-axis as a fraction of circular component
	g	Helical curve elevation in Z-axis
	α	Turning angle around filament tangent for two consecutive monomers
	h_l^I	Length to complete an inner turn of structural units on a center-line curve
	P_b	Branching probability for filament networks
Macromolecule	d	The variable distance between two consecutive macromolecules in a cluster
	$\mathcal{S}$	A sphere in $\mathbb{R}^3$ defined by a center and a radius
	J	Number of different macromolecules in a cluster
Transformation	R	Rotation
	T	Translation
Algorithm	Q	A list (or queue)
	G	Models (structural or transformation) generator
	O	Occupancy
	N	Maximum number of tries

Table 2: Glossary mathematical symbols defined in the main publication.

FILE	VARIABLE	DESCRIPTION	DATA TYPE
Membrane	MB_TYPE	membrane geometry	'sphere', 'ellipse' or 'torus'
.mbs	MB_OCC	O	range [low, min] in percentages
	MB_THICK_RG	t	range in Å
	MB_LAYER_S_RG	σ_l	range in Å
	MB_MAX_ECC	max ϵ	positive real value
	MB_OVER_TOL	overlapping tolerance	percentage
	MB_MIN_RAD	min r	positive real value in Å
	MB_DEN_CF_RG	density factor	range of reals between 0 and 1
Filament	HLIX_TYPE	filament type	'actin' or 'mt'
.hns	HLIX_MMER_RAD	monomer radius r_m	positive real value in Å
	HLIX_PMER_L	d_M/r_m	positive real value
	HLIX_PMER_OCC	O	range in percentages
	HLIX_MIN_P_LEN	min p	positive real value
	HLIX_HP_LEN	h_l^I	positive real value in Å
	HLIX_MZ_LEN	monomer length in z-axis	positive real value in Å
	HLIX_MZ_LEN_F	min h_z	range of reals between 0 and 1
	HLIX_OVER_TOL	overlapping tolerance	percentage
	HLIX_MIN_NMMER	min structural units	positive integer
	A_BPROP	P_b	range reals between 0 and 1
	A_MAX_P_BRANCH	max P_b	range reals between 0 and 1
	MT_RAD	MT radius Z-ring	positive real value in Å
	MT_NUNITS	MT ring number of monomers	positive integer
Mmolecules	MMER_ID	identificador	string
.pns or .pms	MMER_SVOL	path to density (.mrc)	string
	MMER_ISO	isosurface threshold	positive real value
	PMER_L	$d_i \in r_m \cdot$ [PMER_L, 2xPMER_L]	positive value
	PMER_OCC	O	percentage
	PMER_LMEX	cluster max length	positive real value in Å
	PMER_OVER_TOL	overlapping tolerance	percentage
	PMER_REVERSE_NORMALS	reverse membrane normal	boolean

Table 3: Variables used by the input files to model the generators.

TYPE	STRUCTURE	LABEL
Membrane	Spherical	1
	Ellipsoid	2
	Toroid	3
Filament	Microtubules	4
	Actin Network	5
Cytosolic Protein	4v4r	6
	3j9i	7
	4v4r_50S	8
	4v4r_30S	9
	6utj	10
	5mrc	11
	4v94	12
	4cr2	13
	3qm1	14
	3h54	15
	3gl1	16
	3d2f	17
	3cf3	18
	2cg9	19
	1u6q	20
	1s3x	21
	1qvr	22
	1bxn	23
Membrane protein	5wek	24
	4pe5	25
	5ide	26
	5gjv	27
	5kxi	28
	5tj6	29
	5tqq	30
	5vai	31

Table 4: Table with the structures used to simulate the cellular context. Proteins are identified by their Protein Data Bank (PDB) code. 4v4r_50S and 4v4r_30S stand for the isolated subunits 50S and 30S respectively of 4v4r complex. The label is the integer assigned to the ground truth output data.