

Supplementary Information

High-throughput robotic collection, imaging, and machine learning analysis of salt patterns: Composition and concentration from dried droplet photos

Bruno C. Batista^a, Amrutha S. V.^a, Jie Yan^b, Beni B. Dangi^c and Oliver Steinbock^{a,*}

^a Florida State University, Department of Chemistry and Biochemistry, Tallahassee, FL 32306, USA;

^b Bowie State University, Department of Computer Science, Bowie, MD 20715, USA;

^c Florida Agricultural and Mechanical University, Department of Chemistry, Tallahassee, FL 32307, USA.

* Email: osteinbock@fsu.edu

Movie 1: Time-lapse movie of RODI operation, specifically the spaced placement of solution drops. Filename: movie1.mp4

Movie 2: Sequence of deposit patterns for which our trained AI system identifies composition and initial concentration. The photos were not included in the training set. Filename: movie2.mp4

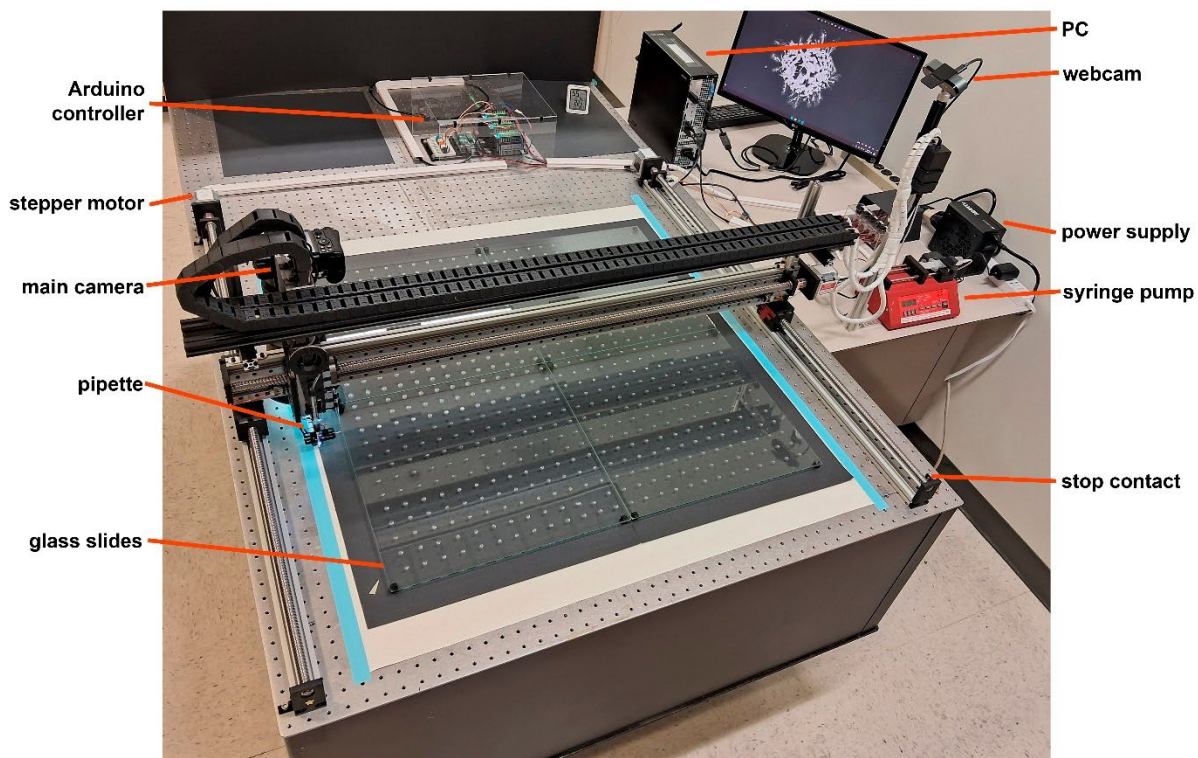


Fig. S1. Photo of the robotic drop imager RODI with labeled components.

Table S1. Compilation of RODI parts and suppliers.

	Part	Supplier
1	Dell Optiplex 7010	Dell
2	Corsair RM1000e (2023) Fully Modular Low-Noise Power Supply	Corsair Store (Amazon)
3	Arduino Student Kit [AKX00025]	Arduino Store (Amazon)
4	Stepper Motor Drivers, 4 Pack 4A Stepper Driver Controller CNC Digital Micro Step Driver for 39/42/57 Stepper Motor	MECCANIXITY (Amazon)
5	FUYU FSK40 Gantry Linear Guide Stage XYZ Slide Table X=1000mm,Y=1000mm,Z=50mm Stroke Work Size Linear Stage Module for DIY	FUYU Store (Amazon)
6	Nikon Z 5 mirrorless camera, USA model	Nikon Store (Amazon)
7	Nikon NIKKOR Z MC 105mm f/2.8 VR S Professional macro prime lens, USA model	Nikon Store (Amazon)
8	62mm MC UV Protection Filter Slim Frame with 18-Multi-Layer Coatings for Camera Lens (K-Series)	K&F Concept Store (Amazon)
9	Micro Center 128GB Class 10 MicroSDXC Flash Memory Card with Adapter	INLAND Store (Amazon)
10	White LED Strip Lights, 16.4ft Dimmable Bright Rope Light, 6000K Bright Daylight White Tape Light, 5M 12V Ribbon Light, 300 LEDs Flexible	Onforu (Amazon)
11	Custom glass cut to size	HOME4 Store (Amazon)
12	Amazon Basics Power Extension Cord, 25 Feet, 13 Amps, 125V, Black	Amazon Basics Store
13	6 Ft Power Strip Surge Protector - 7 Outlets 4 USB Ports (2 USB C), Maxpw Ultra Thin Flat Extension Cord & Flat Plug, 1700 Joules, Wall Mount, Desk Charging Station for Home Office Dorm, White	Winpw (Amazon)

14	FEP Teflon High Temperature Wire of 30awg with Red Blue Green Yellow Black colors 10m Each in Box Magnet Wire	Striveday (Amazon)
15	FEP Teflon High Temperature Wire of 20awg with Red Blue Green Yellow Black Colors in Box Magnet Wire. Each Color 8m	Striveday (Amazon)
16	(Real 18AWG) 10 Pairs 12V 5A DC Power Pigtail Barrel Plug Connector Cable, 2.1mm x 5.5mm Male Female DC Pigtail Connectors for CCTV Security Camera and 12V Power Supply Adapter	MILAPEAK (Amazon)
17	40ft 18 AWG Gauge Electrical Wire, Premium DC 12V Hookup Red Black Copper Stranded Auto 2 Cord, Flexible Extension Cable with Spool	MILAPEAK (Amazon)
18	10pcs Micro Mechanical Switch 3D Printer Part End Stop Limit Switch 3 Pin Compatible with CNC RAMPS 1.4 RepRap 3D Printer CR-10 10S,S4,S5,Ender 3/Ender 3 Pro/Ender 3 V2	AEDIKO US (Amazon)
19	DaierTek ON Off Round Rocker Switches SPST 20mm Mini 12 Volt DC Mini Circle Toggle Switch 12V for Car Automotive RV 2 Pin Switch 120V Wired KCD1-5Pack	DaierTek (Amazon)
20	HATCHBOX 1.75mm Black PLA 3D Printer Filament, 1 KG Spool, Dimensional Accuracy +/- 0.03 mm, 3D Printing Filament	HATCHBOX Store (Amazon)

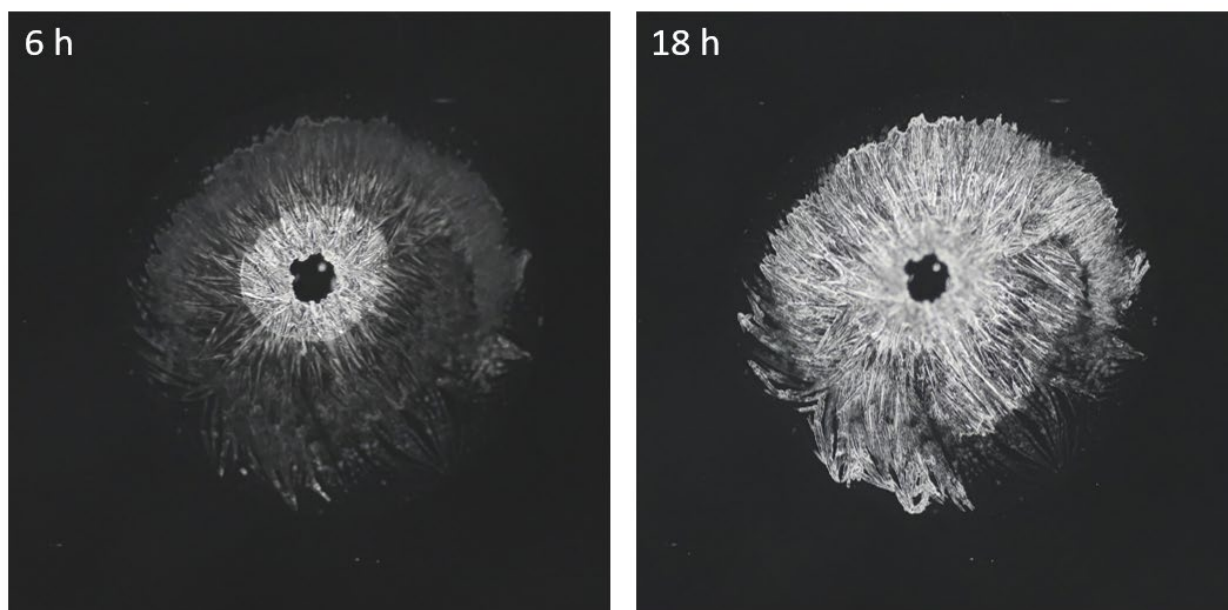


Fig. S2. Time evolution of the deposit pattern formed during the drying of a 50% v/v Na_2SO_3 sessile drop. The panels show the evaporite after 6 and 18 h of experiment. Field of view $1 \times 1 \text{ cm}^2$.

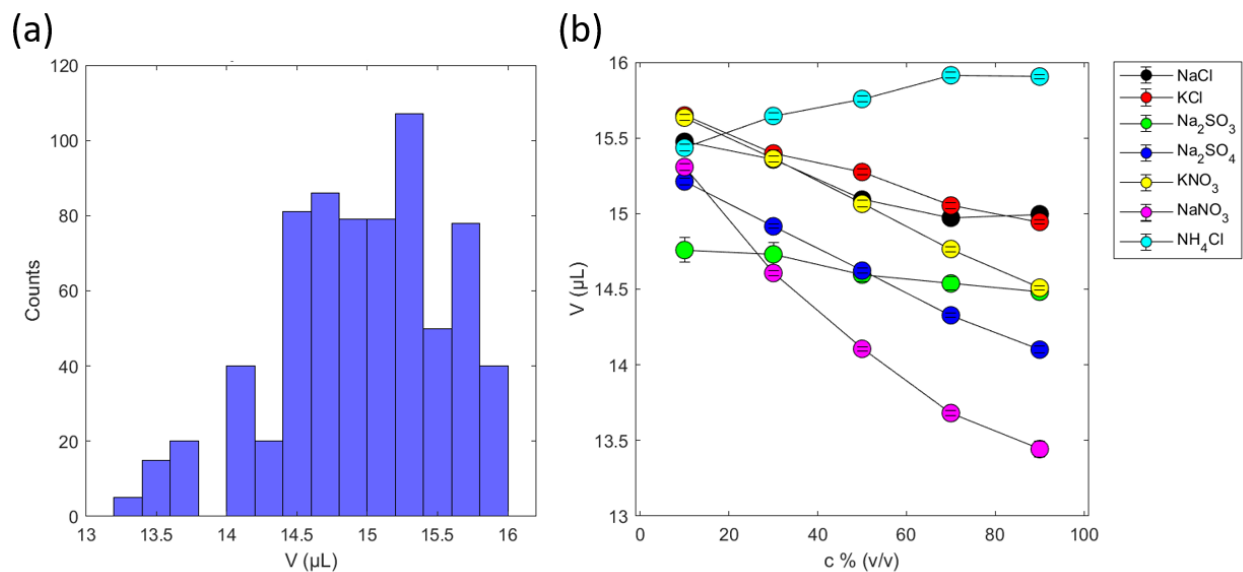


Fig. S3. Volume of dispensed drops as a function of salt concentration and composition. (a) Histogram of the total 700 volume data points originating from seven salts, five concentrations, and 20 repeats each. The mean value and standard deviation are $15.0 \pm 0.6 \mu\text{L}$. (b) Volume dependence on concentration for the salts listed in the legend. Each point corresponds to the average value of 20 repeats.

Table S2. Number of images excluded from each of the 35 categories (7 salts and 5 concentrations). The total number of excluded images is 206 out of 23,660 (i.e. 0.87 %).

Salt	Concentrations				
	10	30	50	70	90
NH ₄ Cl	13	11	11	3	3
NaCl	3	1	1	4	11
KCl	5	3	16	1	5
Na ₂ SO ₄	9	5	5	5	12
KNO ₃	12	7	16	1	14
Na ₂ SO ₃	1	9	1	0	1
NaNO ₃	11	6	8	4	2

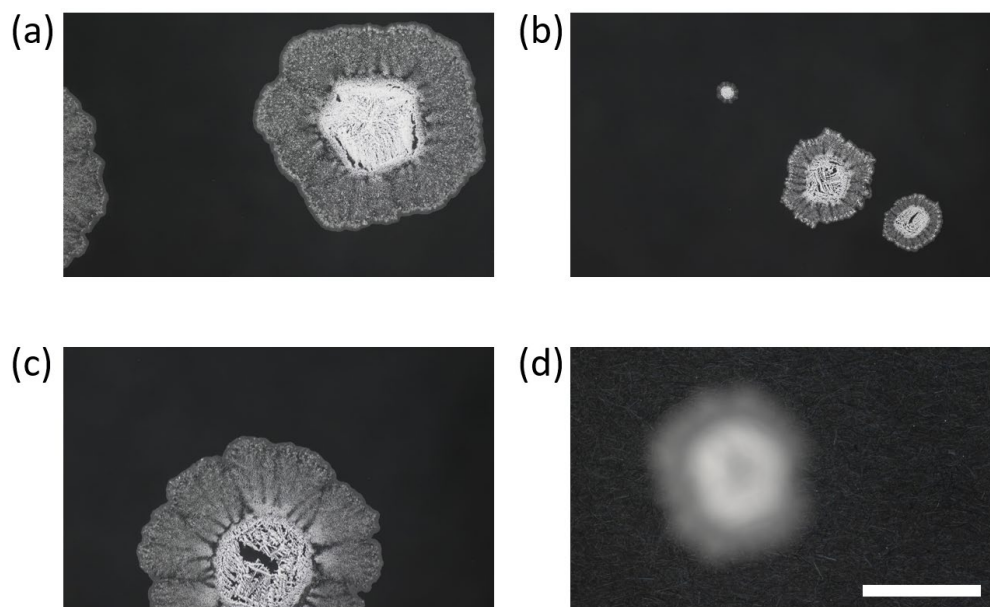


Fig. S4. Examples of excluded images. (a) Image captured a part of the neighboring deposit pattern, (b) Splashing resulted in the formation of many small deposit patterns, (c) The deposit pattern is not captured completely, (d) Out of focus. Scale bar = 1 cm

Table S3. Saturation concentration of different aqueous salt solutions as taken from D. R. Lide, Ed., CRC Handbook of Chemistry and Physics, Vol. 85. CRC press, 2004.

Salt	Solubility (g/L) at 25°C
NH ₄ Cl	395
NaCl	360
NaNO ₃	912
Na ₂ SO ₃	307
Na ₂ SO ₄	281
KCl	355
KNO ₃	383

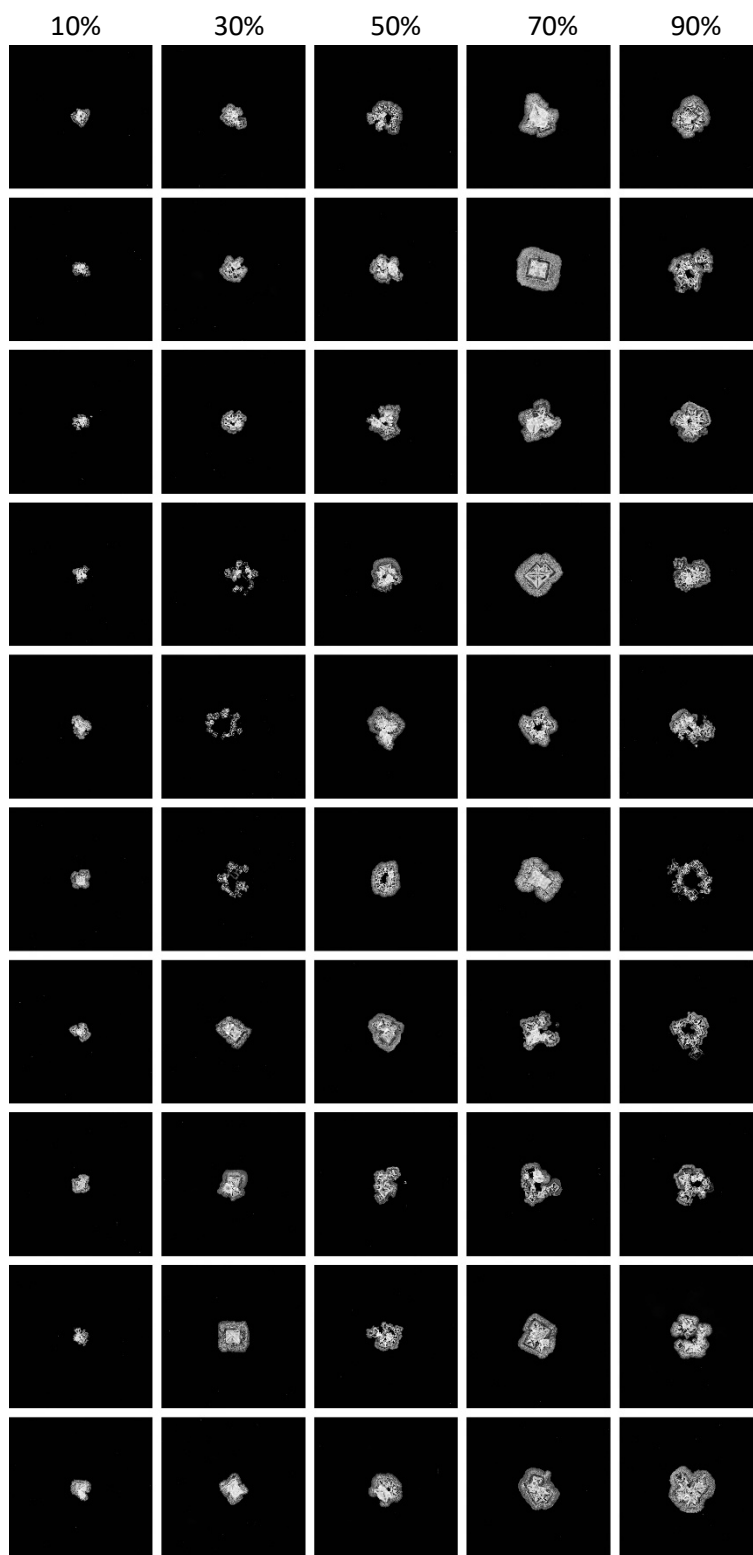


Fig. S5. Deposit patterns formed by aqueous NaCl solutions. Each of the five columns corresponds to vol/vol concentrations of 10%, 30%, 50%, 70%, and 90%, where 100% would denote the saturated solution. Area shown in each individual image: $2.0 \times 2.0 \text{ cm}^2$.

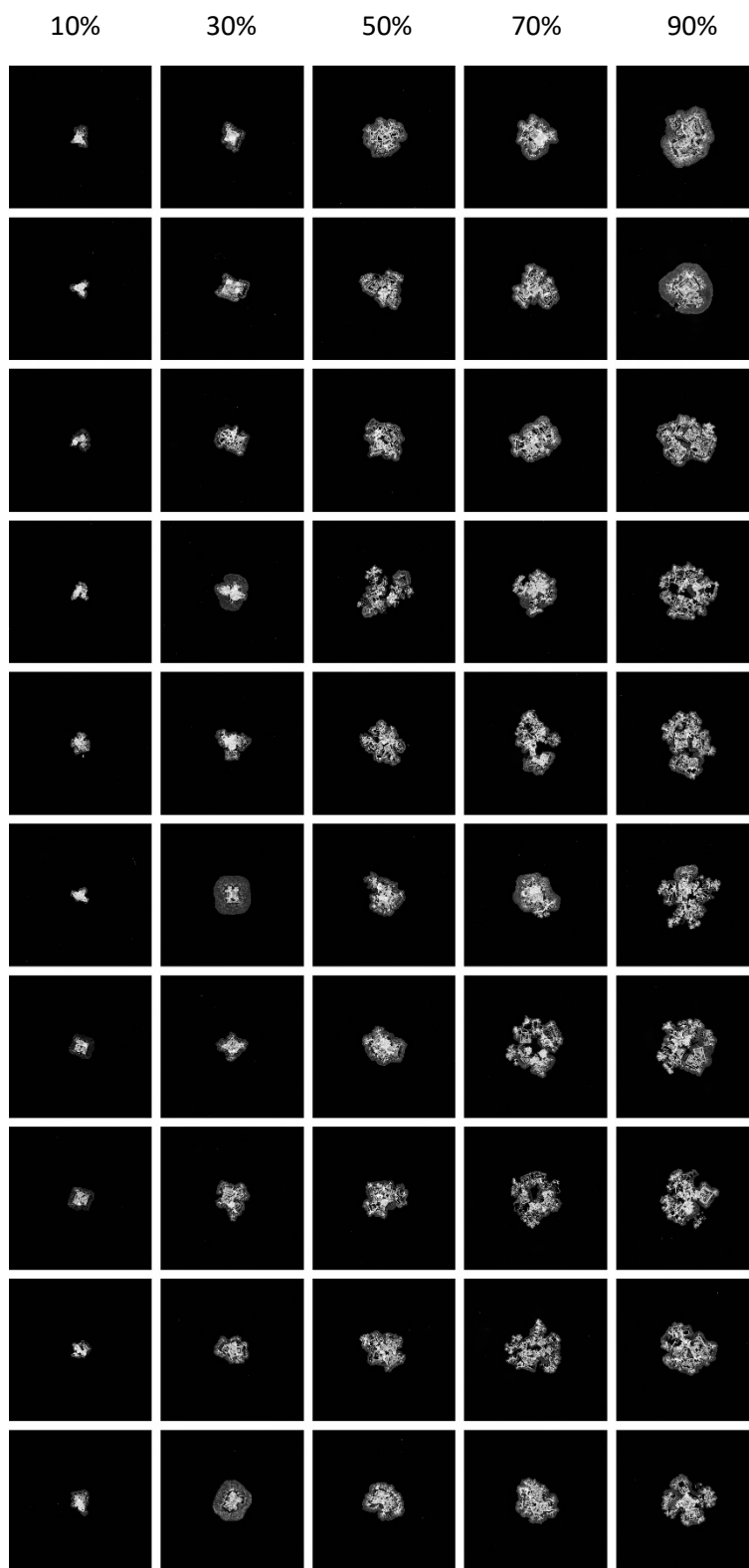


Fig. S6. Deposit patterns formed by aqueous KCl solutions. Each of the five columns corresponds to vol/vol concentrations of 10%, 30%, 50%, 70%, and 90%, where 100% would denote the saturated solution. Area shown in each individual image: $2.0 \times 2.0 \text{ cm}^2$.

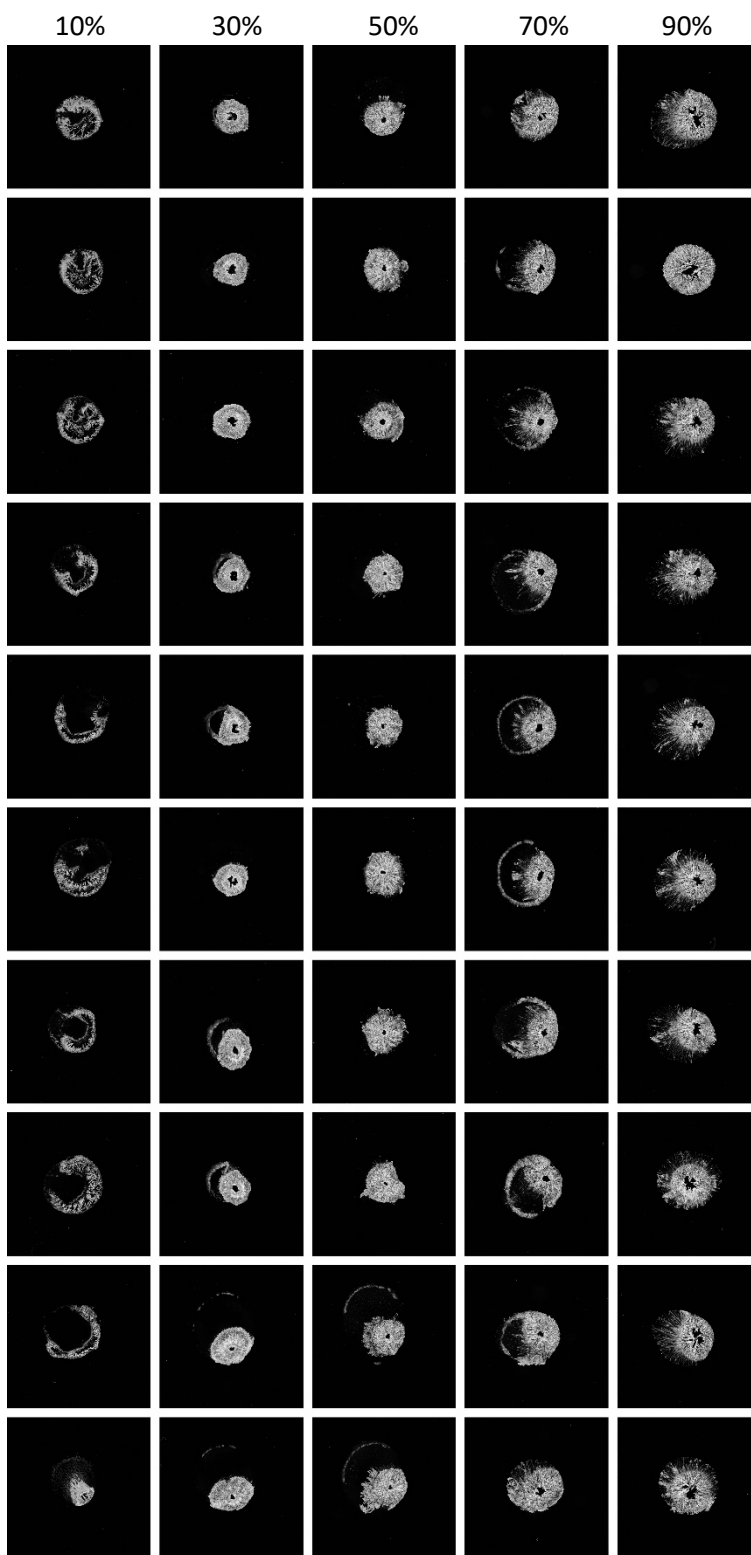


Fig. S7. Deposit patterns formed by aqueous Na_2SO_3 solutions. Each of the five columns corresponds to vol/vol concentrations of 10%, 30%, 50%, 70%, and 90%, where 100% would denote the saturated solution. Area shown in each individual image: $2.0 \times 2.0 \text{ cm}^2$.

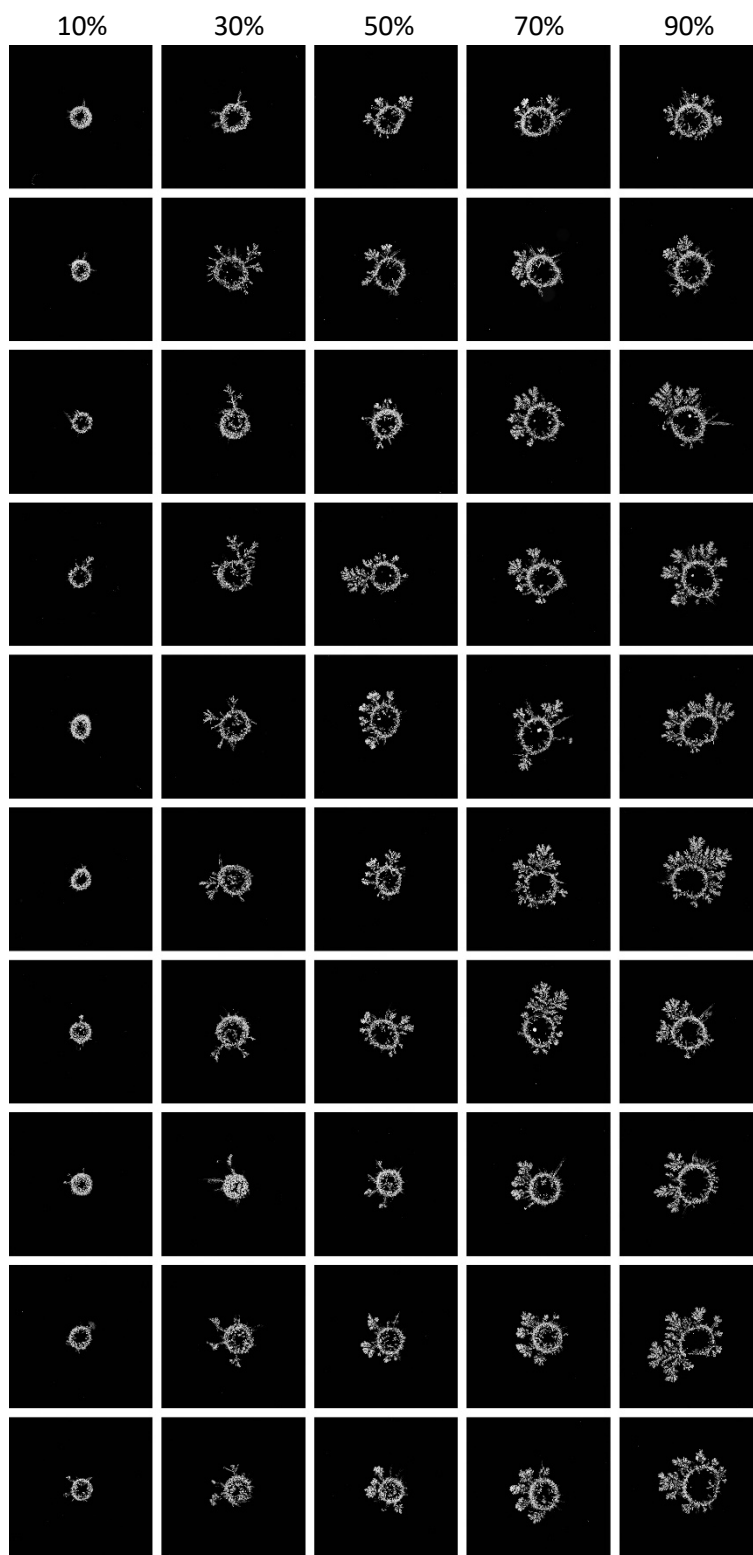


Fig. S8. Deposit patterns formed by aqueous Na_2SO_4 solutions. Each of the five columns corresponds to vol/vol concentrations of 10%, 30%, 50%, 70%, and 90%, where 100% would denote the saturated solution. Area shown in each individual image: $2.0 \times 2.0 \text{ cm}^2$.

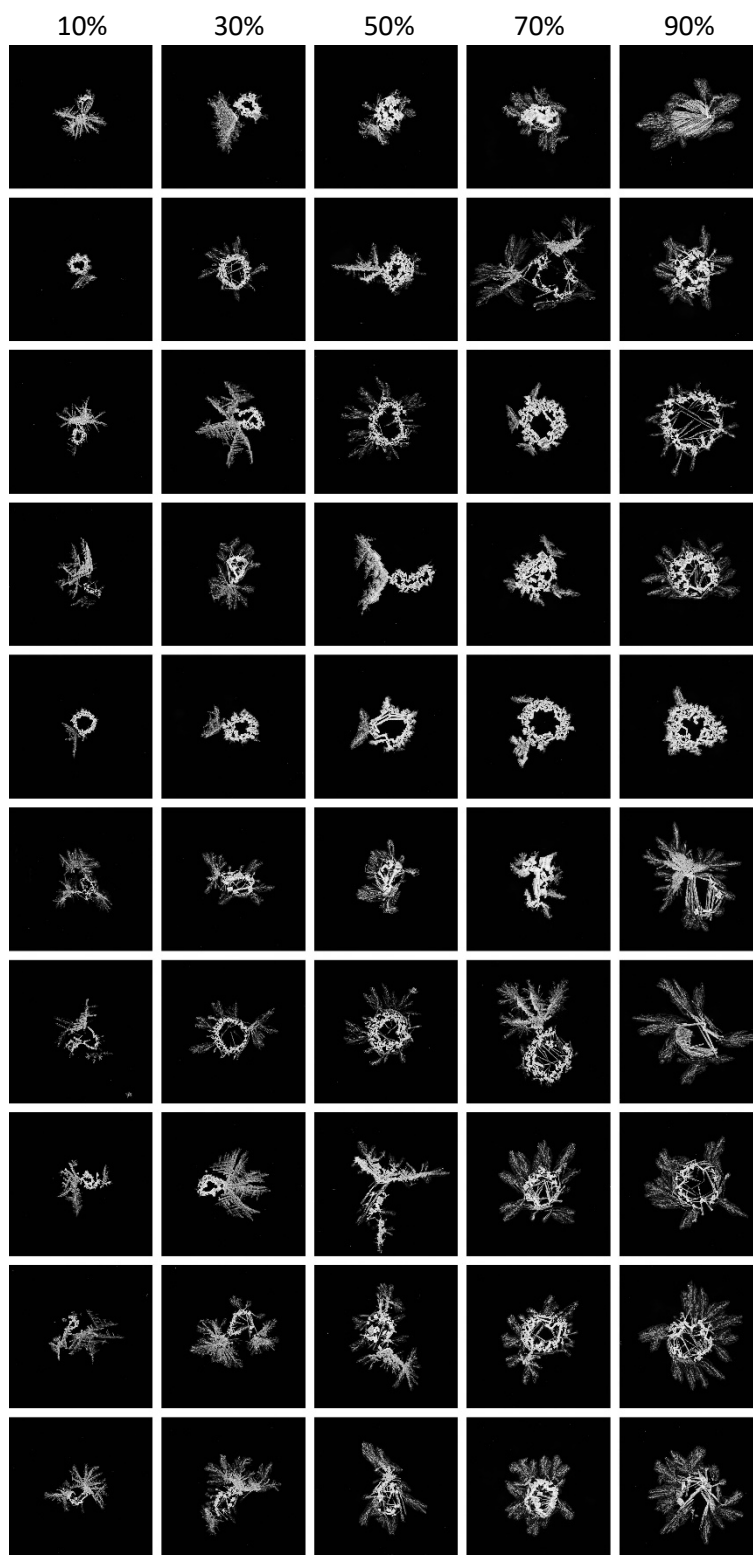


Fig. S9. Deposit patterns formed by aqueous KNO_3 solutions. Each of the five columns corresponds to vol/vol concentrations of 10%, 30%, 50%, 70%, and 90%, where 100% would denote the saturated solution. Area shown in each individual image: $2.0 \times 2.0 \text{ cm}^2$.

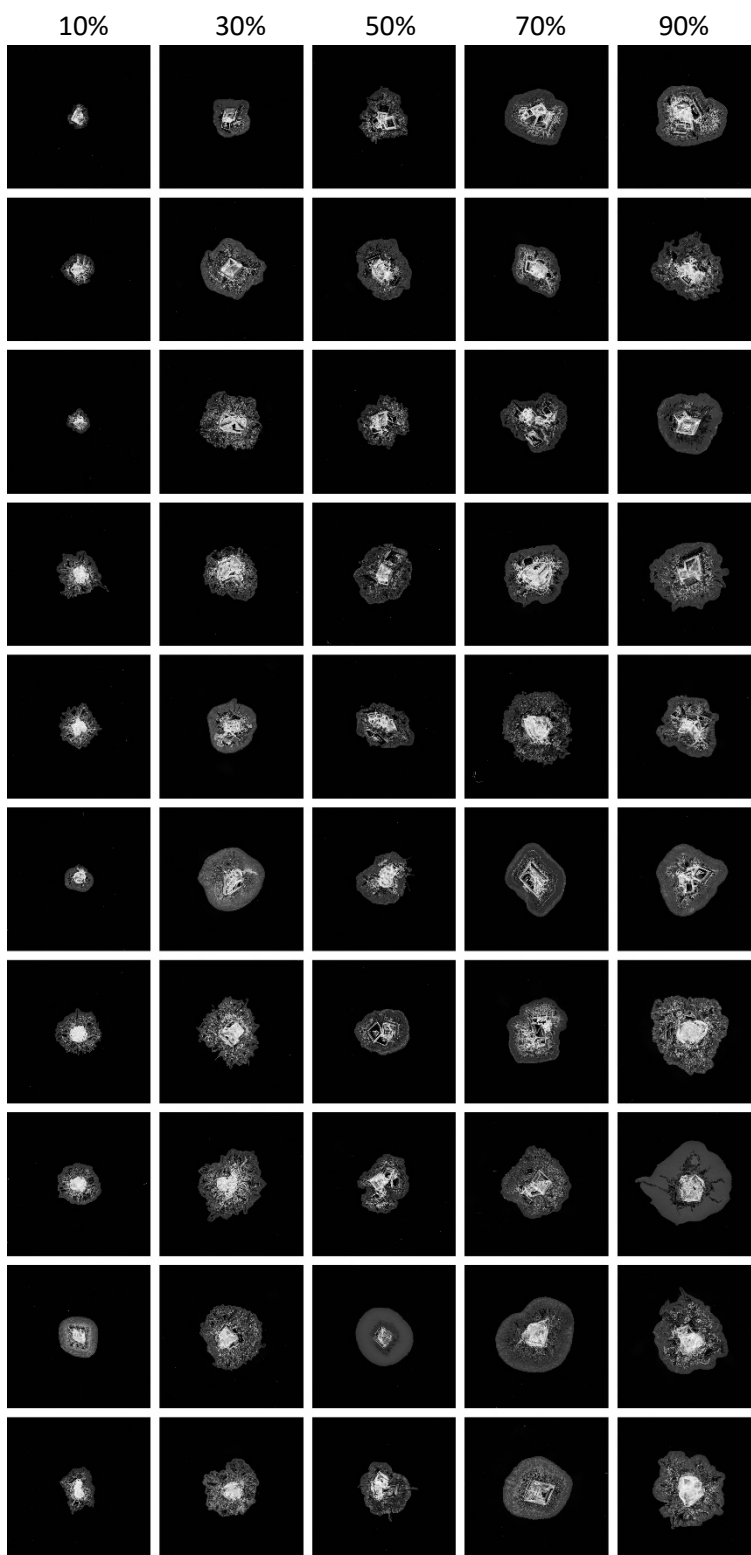


Fig. S10. Deposit patterns formed by aqueous NaNO_3 solutions. Each of the five columns corresponds to vol/vol concentrations of 10%, 30%, 50%, 70%, and 90%, where 100% would denote the saturated solution. Area shown in each individual image: $2.0 \times 2.0 \text{ cm}^2$.

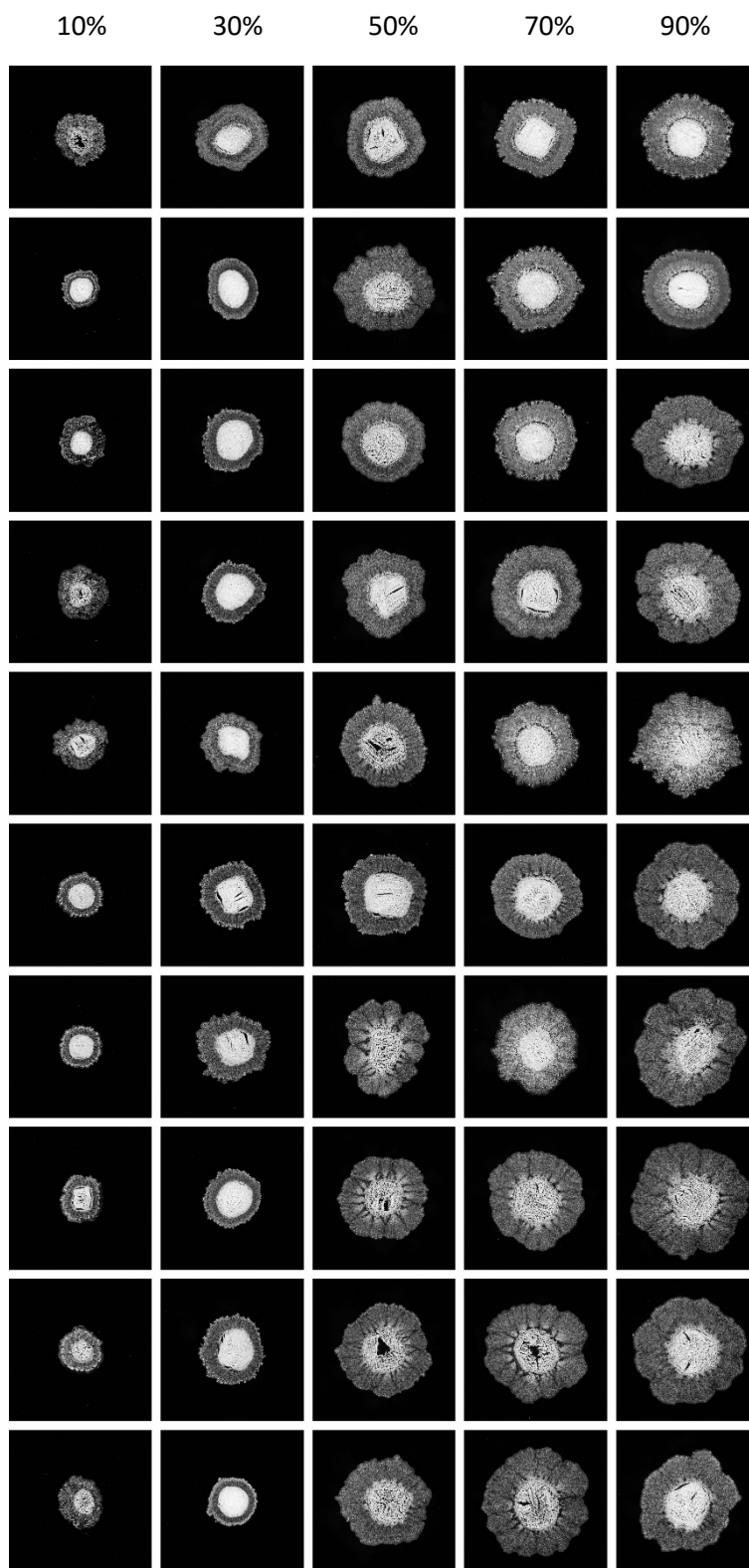


Fig. S11. Deposit patterns formed by aqueous NH_4Cl solutions. Each of the five columns corresponds to vol/vol concentrations of 10%, 30%, 50%, 70%, and 90%, where 100% would denote the saturated solution. Area shown in each individual image: $2.0 \times 2.0 \text{ cm}^2$.

Table S4. Summary of all 47 metrics with names and short descriptions. The calculation of these metrics involves two intensity thresholds (low 70, high 200). All images analyzed have the same spatial resolution and size, where 1 cm = 1620 pixels (and equivalently 1 pixel = 6.18 μm). The values of `maxLargeHoleAreas`, `medianLargeHoleAreas` are set to zero if no holes are found. The high-threshold metrics `medianEccentricity` and `medianArea` are set to the category mean if no large connected areas are found.

1	numWhitePixels	This measure is the total count of white pixels. It specifies the total deposit area.
2	numBlackPixels	This measure is the total count of black pixels within regions surrounded by white pixels. Notice that this quantity is sensitive to small gaps in the white regions that connect the black region to the global background. If such a gap exists, the black area is not analyzed.
3	ratio	This measure is the ratio of the pixel counts in 2 and 1.
4	numLargeBlobs	This measure is the total number of connected white areas.
5	perimeterLength	This measure is the sum of the perimeter lengths of all connected white areas. For a given total white area, it increases with <code>numLargeBlobs</code> and the eccentricity of the individual blobs.
6	axisRatio	This measure is the eccentricity as calculated from the best-fit ellipse for all white pixels. Values larger than one indicate that the deposit deviates from a circular disk.
7	countLargeHoles	This measure is the number of black connected areas (holes) larger than 1000 pixels.
8	medianLargeHoleAreas	This measure is the median value of the black connected areas (holes) larger than 1000 pixels.
9	maxLargeHoleAreas	This measure is the maximum value of the black connected areas (holes) larger than 1000 pixels.
10	meanDistances	This measure is the average of the distances of all white pixels from their common centroid.
11	stdDistances	This measure is the standard deviation of the distances of all white pixels from their common centroid.
12	modeDistances	This measure is the most frequent value among the distances of all white pixels from their common centroid.

13	medianDistances	This measure is the median of the distances of all white pixels from their common centroid.
14	skewnessDistances	This measure is the degree of asymmetry observed in the distribution of the distances of all white pixels from their common centroid. Zero implies a symmetric distribution, whereas positive (negative) values indicate that the distribution is skewed to the right (left).
15	erosionslope	We compute the fraction of the remaining white pixels f after erosion with disks of radius r . The slope of $f(r)$ for small disk radii (0-4 pixels) defines this measure. Large values indicate the presence of fine details in the deposit pattern.
16	frct01	We compute the fraction of the remaining white pixels f after erosion with disks of radius r . The smallest disk radius for which $f(r) \leq 0.1$ defines this integer measure. Large values indicate compact deposit patterns such as featureless white disks.
17	medianEccentricity	The median eccentricity of connected regions above the high threshold. Eccentricity measures how elongated a shape is, with values closer to 1 indicating more elongated shapes.
18	medianArea	The median area of connected regions above the high threshold. This represents the typical size of the bright precipitate regions.
19	sumEdgesLow	The ratio of the edge points detected in the low-threshold binary image to the precipitate area. It provides a measure of edge density, indicating how jagged or smooth the precipitate boundary is in less intense regions.
20	sumEdgesHigh	The ratio of the edge points detected in the high-threshold binary image to the precipitate area. This metric focuses on the density of edges in the brighter, more intense regions of the precipitate.
21	areaOverEdgeLow	The ratio of the total precipitate area to the number of edge points in the low-threshold image. A higher value suggests larger, more contiguous precipitate regions relative to their boundary length.
22	areaOverEdgeHigh	The ratio of the total precipitate area to the number of edge points in the high-threshold image. This metric assesses the relationship between the area of brighter regions and their boundary complexity.

23	stdRaw	The standard deviation of pixel intensities above the low threshold. It quantifies the variability in intensity within the precipitate, indicating how uniform or heterogeneous the precipitate is.
24	areaHigh	The total number of pixels above the high threshold, representing the area of the more intense, bright precipitate regions.
25	stdHigh	The standard deviation of pixel intensities above the high threshold. This measures the intensity variability within the bright regions of the precipitate.
26	compactnessCenter	The fraction of pixels above the low threshold within a defined central disk (radius radiCenter = 200 pixels) around the centroid. This indicates how densely packed the precipitate is in the core region.
27	brightnessCenter	The average brightness of pixels within the central disk (radius radiCenter = 200 pixels) around the centroid. This metric provides an overall measure of the intensity in the core region.
28	blackCoreFraction	The fraction of pixels below the low threshold within the central disk, indicating the proportion of dark areas in the core region relative to the total core area.
29	intensityKurtosis	The kurtosis of pixel intensities above the low threshold. Kurtosis measures the "tailedness" of the intensity distribution, with higher values indicating more pronounced peaks.
30	intensitySkewness	The skewness of pixel intensities above the low threshold. Skewness measures the asymmetry of the intensity distribution, with positive values indicating a right-skewed distribution and negative values indicating a left-skewed distribution.
31	intensityRatio	The ratio of average intensities between inner and outer ring-sections of the precipitate. This metric assesses radial intensity variation from the center outwards.
32	skeletonLength	The ratio of the length of the skeletonized precipitate (a representation of its structure) to the precipitate area. This indicates structural complexity and connectivity within the precipitate.

33	skeletonBranchPoints	The ratio of skeleton branch points (junctions in the skeleton) to the precipitate area, indicating the complexity and branching nature of the structure.
34	skeletonEndPoints	The ratio of skeleton endpoints to the precipitate area, providing a measure of the number of terminal points in the skeleton.
35	fractalDim	An estimate of the fractal dimension, representing the complexity and self-similarity of the precipitate structure. Higher values indicate more complex, self-similar structures.
36	log10Entropy	The log-transformed entropy of the image normalized by the precipitate area. Entropy measures the randomness of pixel intensities, with higher values indicating more complex textures.
37	waveletEntropy	The entropy of wavelet coefficients normalized by the precipitate area, representing the complexity and variability of textures at different scales.
38	stdRays	The standard deviation of average intensities along radial directions from the center. This measures how much the intensity varies as you move outwards in different directions.
39	lowRays	This metric measures the median intensity of the darkest 10% of radial lines extending from the center of the precipitate outward. It evaluates the average pixel intensity along each radial line and focuses on the dimmest regions to capture variations in brightness across different directions. This metric helps quantify the spread of low-intensity areas within the precipitate, providing insights into uneven material or light distribution.
40	stdMaxRays	The standard deviation of the largest radii found along different angles from the center, indicating the variation in the boundary distance from the center.
41	corrGLCM	The correlation from the Gray-Level Co-Occurrence Matrix (GLCM), which measures the relationship between pixel intensities and their spatial dependencies, indicating texture consistency.
42	energyGLCM	The energy from the GLCM, which quantifies the uniformity of textures. Higher values indicate more homogeneous textures.

43	meanStd5	The average local standard deviation (calculated with a disk radius of 5 pixels) normalized by the precipitate area, indicating small-scale textural variation.
44	meanStd25	The average local standard deviation (calculated with a disk radius of 25 pixels) normalized by the precipitate area, indicating medium-scale textural variation.
45	ms25over5	The ratio of medium-scale to small-scale local standard deviations, providing insight into the relative textural variation at different scales.
46	ms100over25	The ratio of large-scale (disk radius of 100 pixels) to medium-scale local standard deviations, further indicating textural variation at different levels.
47	numContours	The ratio of the number of contours detected in the high-threshold image to the precipitate area, indicating the complexity and number of distinct regions within the precipitate.

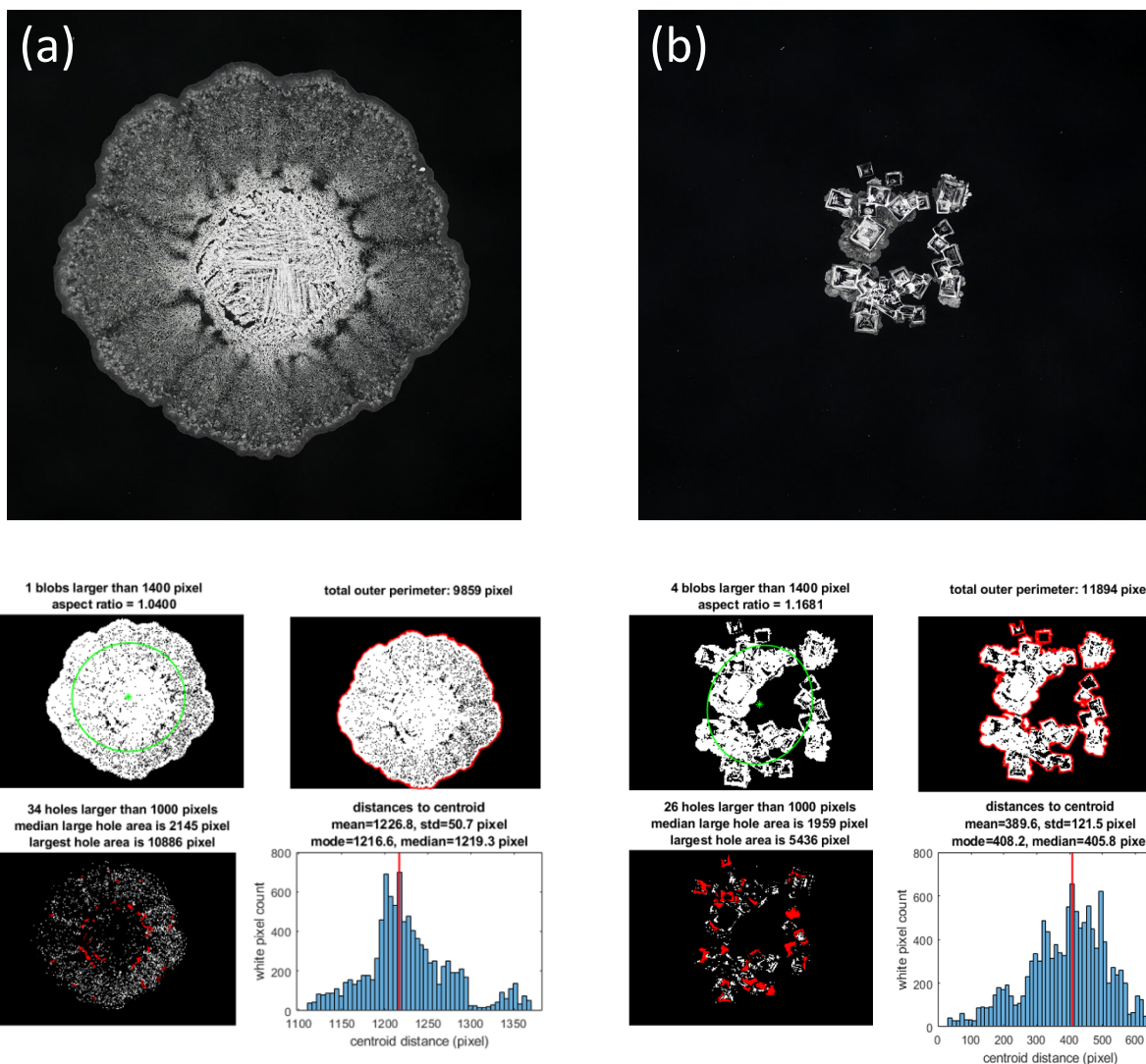


Fig. S12. Examples of deposit patterns formed by (a) NH_4Cl (90%) and (b) NaCl (90%). Field of view: $2 \times 2 \text{ cm}^2$. The smaller panels underneath illustrate some of the analyses that were performed to determine the image metrics. These include eccentricity (green ellipse), perimeter length, dark holes, and white pixel distances from the pattern centroid. The names of the corresponding metrics are: `axisRatio`, `perimeterLength`, `countLargeHoles`, `maxLargeHoleAreas`, `meanDistances`, `stdDistances`, `modeDistances`, and `medianDistances`. Table S3 shows the values of all 47 metrics for these two example patterns.

Table S5. Values of all 47 metrics for the patterns in Fig. S12 before Z-scoring. For averaged Z-scored results see Fig. 6 in the main paper.

No.	Name of metric	NH ₄ Cl example in (a)	NaCl example in (b)
1	numWhitePixels	4141767	399633
2	numBlackPixels	90815	58605
3	ratio	0.021926632	0.146647049
4	numLargeBlobs	1	4
5	perimeterLength	9859	11894
6	axisRatio	1.040014781	1.168094094
7	countLargeHoles	34	26
8	medianLargeHoleAreas	2145	1959
9	maxLargeHoleAreas	10886	5436
10	meanDistances	1226.752525	389.5504176
11	stdDistances	50.66848028	121.5245205
12	modeDistances	1216.6	408.2
13	medianDistances	1219.263601	405.7956535
14	skewnessDistances	0.523409967	-0.47927002
15	erosionslope	0.092823498	0.089701801
16	frct01	17	14
17	medianEccentricity	0.822599941	0.886296169
18	medianArea	775	897
19	sumEdgesLow	0.101483291	0.133181183
20	sumEdgesHigh	0.053137055	0.080836879
21	areaOverEdgeLow	9.853838867	7.50856826
22	areaOverEdgeHigh	18.81925897	12.37059146
23	stdRaw	47.64925698	47.91466007
24	areaHigh	529179	59363
25	stdHigh	13.95079902	11.17902762
26	compactnessCenter	0.970312239	0
27	brightnessCenter	197.4180801	31.74337957

28	blackCoreFraction	0.000145139	0.031502632
29	intensityKurtosis	2.818117739	1.881786808
30	intensitySkewness	0.911771895	0.372545611
31	intensityRatio	1.381341643	1.040053638
32	skeletonLength	0.131453438	0.123746849
33	skeletonBranchPoints	0.017027731	0.011933251
34	skeletonEndPoints	0.003570886	0.005406523
35	fractalDim	1.997503855	1.9501036
36	log10Entropy	-3.04E-06	-1.67E-05
37	waveletEntropy	5.36E-06	5.53E-05
38	stdRays	8.125542694	15.30274472
39	lowRays	140.6685113	109.5274474
40	stdMaxRays	52.75416088	84.17744307
41	corrGLCM	0.966736432	0.927724104
42	energyGLCM	0.490251318	0.751445395
43	meanStd5	1.25E-06	3.53E-06
44	meanStd25	1.70E-06	5.41E-06
45	ms25over5	1.362874193	1.533723317
46	ms100over25	1.204479187	1.413783915
47	numContours	0.00207675	0.00332861

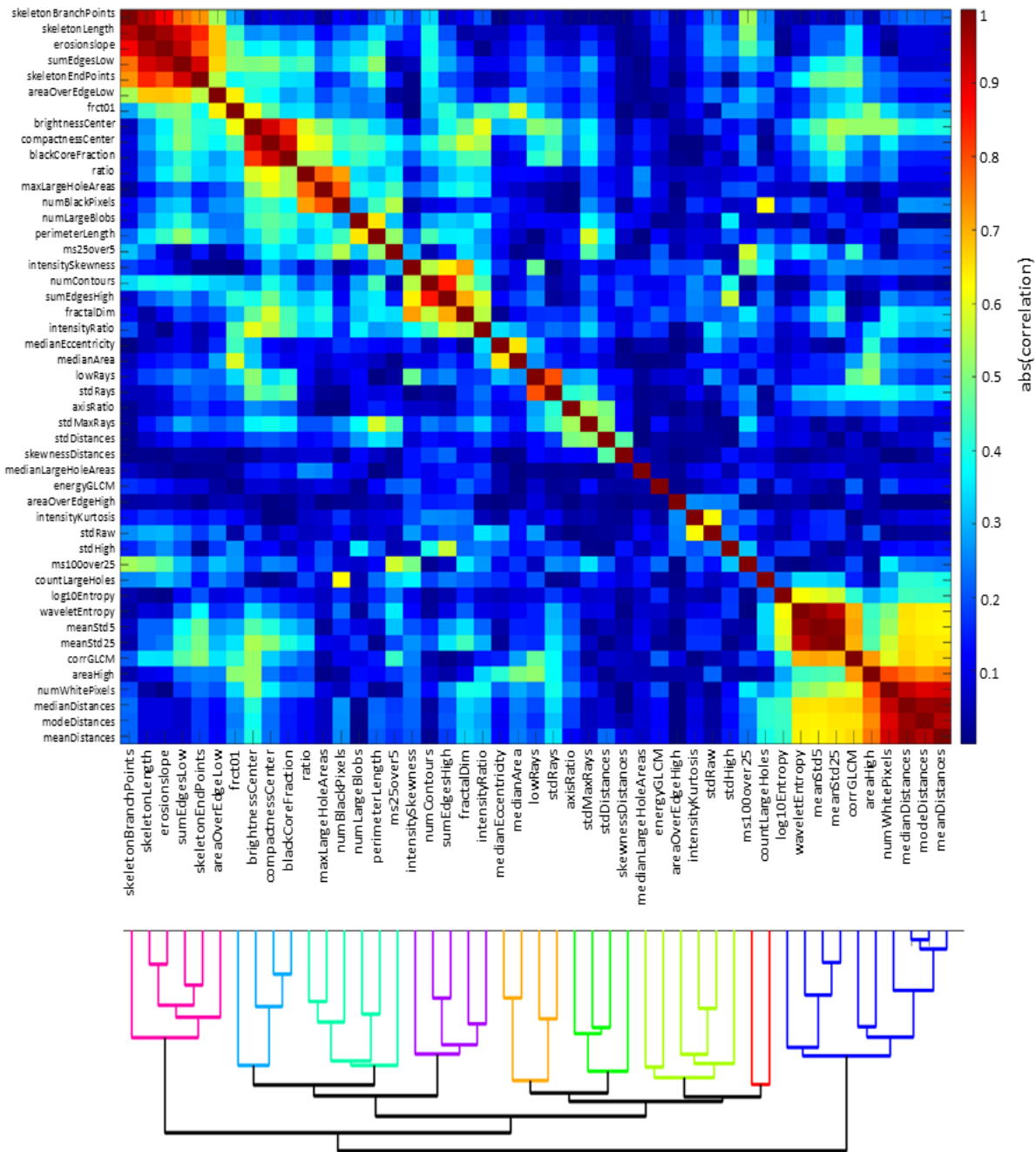


Fig. S13. Correlation analysis of the metrics based on the Z-scored data similar to Fig. 5a but with the hierarchical sorting pattern from Fig. 5b which is partly reprinted underneath the heatmap. The color scale in the heatmap represents the absolute value of the correlation coefficients. The colors in the dendrogram highlight groups of related metrics.

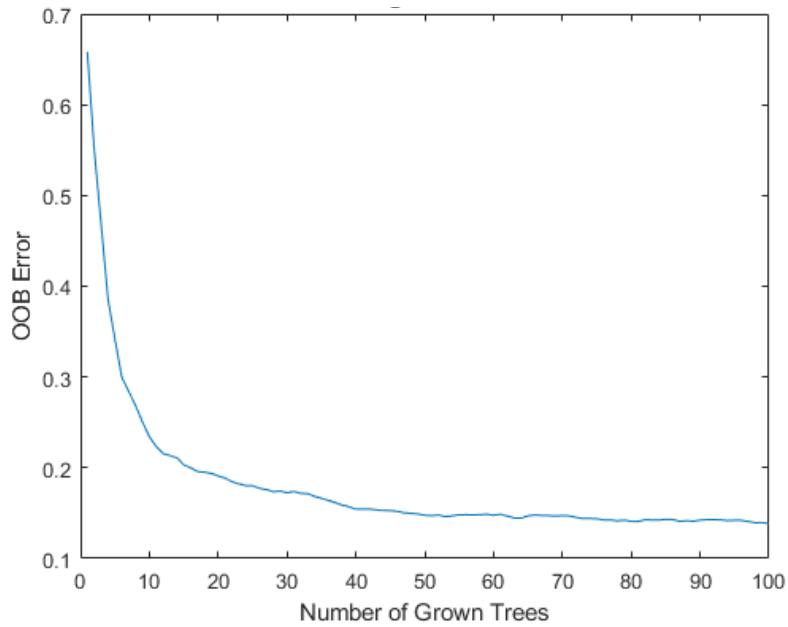


Fig. S14. Out-of-Bag (OOB) error estimate versus the number of grown trees in the Random Forest model. The plot shows the OOB error as a function of the number of trees. Initially, the OOB error is high, indicating underfitting. As the number of trees increases, the OOB error decreases rapidly, demonstrating improved predictive performance. Beyond approximately 60-70 trees, the OOB error plateaus, indicating that additional trees do not significantly enhance accuracy. The final stabilized low OOB error indicates robust generalization to new data. We conclude that a Random Forest model with around 60-70 trees is optimal, and there is no danger of overfitting with 100 trees.

True Class	KCl_10	KCl_30	KCl_50	KCl_70	KCl_90	KNO3_10	KNO3_30	KNO3_50	KNO3_70	KNO3_90	NH4Cl_10	NH4Cl_30	NH4Cl_50	NH4Cl_70	NH4Cl_90	Na2SO3_10	Na2SO3_30	Na2SO3_50	Na2SO3_70	Na2SO3_90	Na2SO4_10	Na2SO4_30	Na2SO4_50	Na2SO4_70	Na2SO4_90	NaCl_10	NaCl_30	NaCl_50	NaCl_70	NaCl_90	NaNO3_10	NaNO3_30	NaNO3_50	NaNO3_70	NaNO3_90		
KCl_10	198	1				1																															
KCl_30		193	6	1			1																														
KCl_50			3	176	17	4		1	1																												
KCl_70				12	168	18			1																											1	
KCl_90					3	20	170																													2	
KNO3_10	5	1				191														1														3			
KNO3_30		2	2		1		181	2													1												10		2		
KNO3_50				1	2		2	182	12													1														2	
KNO3_70					1			12	174	10												1														3	
KNO3_90								12	188														1														1
NH4Cl_10											199																	1						1			
NH4Cl_30												194	4															1									
NH4Cl_50				1								4	195	2	1																						
NH4Cl_70												4	194	4																							
NH4Cl_90													1	4	196																					1	
Na2SO3_10			1													196						1															
Na2SO3_30													1				1	188	7	2	1																
Na2SO3_50																		6	186	8	1	1															
Na2SO3_70																			1	6	168	24															
Na2SO3_90																			2	1	25	168															
Na2SO4_10							1															1															
Na2SO4_30																																					
Na2SO4_50																																					
Na2SO4_70																																					
Na2SO4_90																																					
NaCl_10	1	2	1			4	7	1																													
NaCl_30				2	3		1	2	2	1																											
NaCl_50																																					
NaCl_70																																					
NaCl_90																																					
NaNO3_10						1																															
NaNO3_30																																			196		
NaNO3_50																																		197	1		
NaNO3_70																																		2	191	7	
NaNO3_90																																		5	185	11	
																																		10	191		

Fig. S16. Confusion matrix for identifying the salt type and the initial concentration. The results follow from the ML classifier “XGBoost” and 20 repeats for different random number seeds. We used 70% (N = 16392) of the 23,417 sample entries for training and 30% (N = 7025) for testing. The numbers in the matrix cells denote the averages per run. As obtained from the mean and standard deviation of the repeated runs, the overall accuracy is $(90.1 \pm 0.3)\%$ which is slightly better than the results obtained by the “Random Forest” method. Surprisingly, the two methods show marked differences between resolving the concentration dependence for NaCl and NaNO₃.

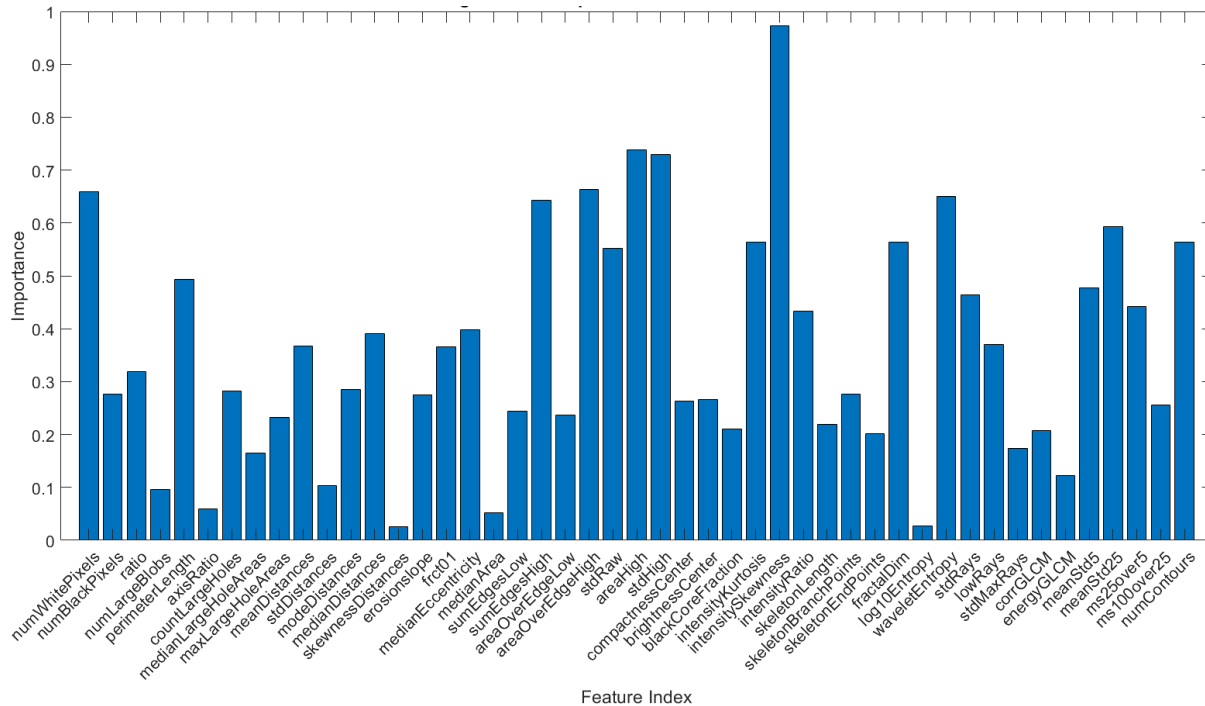


Fig. S17. Feature importance estimates from the Random Forest model. Importance scores are derived using the out-of-bag permutation method, where the values of each feature are permuted, and the impact on the model's accuracy is measured. Higher bars indicate features that have a greater impact on the predictive accuracy of the model. This analysis highlights which features are most critical for making accurate predictions, guiding further feature selection and model refinement. In decreasing order: the highest scores are found for intensitySkewness, areaHigh, stdHigh, areaOverEdgeHigh, numWhitePixels, waveletEntropy, and sumEdgesHigh. The data are the averages of 20 runs and obtained for the prediction of the seven salt categories only.

Execution of MATLAB files:

1. **RODI_MainProcessing.m**: This script processes a sequence of images to analyze dried salt drops. It reads images, applies thresholds for binary conversion, detects and crops edges, and calculates various metrics such as area, eccentricity, and intensity variations. Results are saved to 'results.txt' and renamed manually according to the salt name and concentration.
2. **joinResults.m**: This MATLAB script reads data from the 35 different 'result' files, extracts and processes the salt name and concentration from the file names, adds these as new columns to the data, and combines all the processed data into a single table. The combined table is then optionally saved to a new file. A 'Quality' column is manually added with a default value of 1. A very small number of entries are then set to 0, if the corresponding deposit pattern is strongly atypical (e.g. significant splashing). This process is discussed in the main manuscript.
3. **zScoring.m**: This MATLAB script reads the dataset from the text file saved after executing joinResults.m, excludes rows where the 'Quality' column is zero. It then z-scores all columns except the first nine, creates a new table with the z-scored data, optionally saves it to a new file, and displays a heatmap of the z-scored data.
4. **categoryMeans.m**: This MATLAB script processes the dataset by reading it from the text file containing the z-scored data, calculating averages for specific metric columns based on unique combinations of categories (salts and concentrations) in the data, and displaying the results in two subplots. The upper subplot shows averages by salt type, disregarding concentrations, while the lower subplot shows averages by unique salt-concentration combinations, visualized using heatmaps.
5. **categoryMeans_sorted.m**: This MATLAB script processes data from the tab-delimited text file of the z-scored data, calculates the average of specific columns grouped by unique combinations of values in columns 3 (salt names) and 4 (concentrations), and then creates two sets of averaged data: one by the combination of columns 3 and 4, and another by unique values in column 3. It sorts the data based on standard deviation, displays the results as heatmaps in two subplots, and identifies columns containing NaN values.
6. **correlationsTree.m**: This MATLAB script reads the tab-delimited data file of the z-scored data, computes the correlation matrix of selected columns, and visualizes the correlation matrix in multiple forms: original, reordered by average absolute correlation, and reordered by hierarchical clustering. It also creates a dendrogram to represent the hierarchical clustering of the metrics.
7. **MLP_4e.m**: This MATLAB script reads the tab-separated text file of the z-scored data, extracts relevant columns (salts, concentrations, and metrics), and processes the data for classification. It runs a neural network (Multilayer Perceptron) training process 20 times, each time splitting the data into training and testing sets, handling missing values, normalizing data, and calculating accuracy and

confusion matrices for each run. Finally, it computes the average accuracy, standard deviation, and displays an average confusion matrix.

8. **rf_accuracy.m**: This MATLAB script performs classification of data using Random Forest. It reads a dataset (the z-scored data), splits it into training and test sets (70%/30%), and trains a Random Forest model to predict categories based on given metrics. The script computes accuracies, accumulates confusion matrices, and calculates the mean and standard deviation of accuracy across multiple runs (20). It generates confusion matrices for both cases (salts + concentrations, and salts only), and displays them with the average accuracy.
9. **XGBoost_OnlySalts.m**: This MATLAB script runs the XGBoost model via Python to classify data over 20 repeated runs, calculates and accumulates the average confusion matrix, and computes the average accuracy and standard deviation based on 7 different salt categories. It processes the data (splitting it into training and test sets), trains the model, evaluates its performance, and displays the average confusion matrix with accuracy statistics. The script requires Python 3.8, as indicated by the path to the Python executable.
10. **XGBoost_SaltsAndConcs.m**: This script is similar to the previous one, XGBoost_OnlySalts.m, but performs the training and classification based on 7 different salts as well as 5 different concentrations for each salt.
11. **averageFeatureImportance.m**: This MATLAB script trains a Random Forest classifier on z-scored data, with the goal of evaluating feature importance over multiple runs. It splits the data into training and test sets, trains the model on the training data, and computes the feature importance using out-of-bag error. The script averages the feature importance across multiple runs and then plots the results. It also saves the trained model and the averaged feature importance.
12. **moviePredict1.m**: This MATLAB script processes a dataset from a text file, splits it into training and test sets, and trains a neural network model to predict salt types and concentrations based on given metrics. It handles missing data, normalizes features, and applies categorical labels, while using advanced training options such as early stopping and adaptive learning. After training, the script classifies the test set, selects random test entries, and saves the results (predictions and corresponding information) to a text file.
13. **moviePredict2.m**: This MATLAB script generates a video displaying images along with predicted and true labels (salt type and concentration) for each image based the results saved after executing moviePredict1.m. It processes up to 117 samples, visualizes the results with annotations, and saves the output as a video file (predictionsMLP.mp4). The images are resized and padded to fit a defined figure size, and the script includes conditional color coding for correct or incorrect predictions.

14. **overFittingTest.m**: This MATLAB script trains a Random Forest model on the z-scored data to predict salt types. It splits the data into training and testing sets, trains the model with parameters aimed at reducing overfitting, evaluates the model's performance using out-of-bag error and feature importance, and saves the model and results. While feature importance is analyzed, this script does not perform explicit feature selection based on correlation.

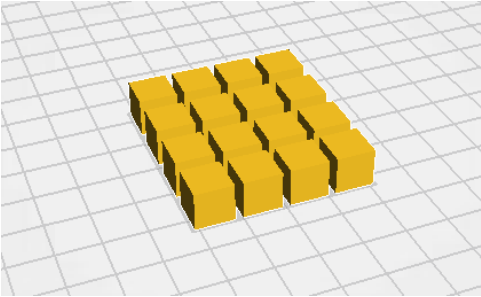
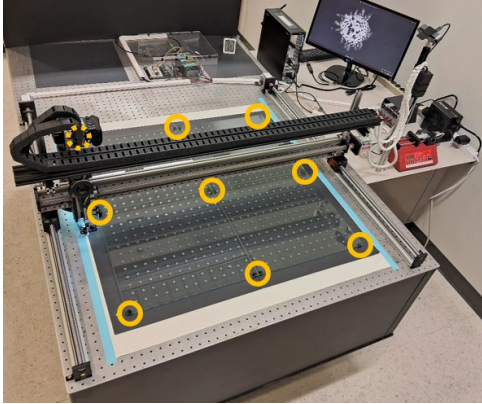
15. **processAllFolders.m**: This MATLAB script loops through a list of specified folders, checks if each folder exists, and attempts to run a corresponding script (named after the folder). If the script exists, it is executed; otherwise, a message is displayed. The script ensures that it returns to the original directory after processing each folder.

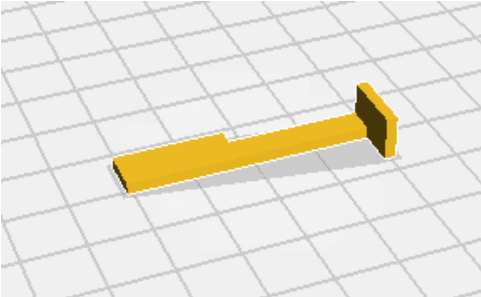
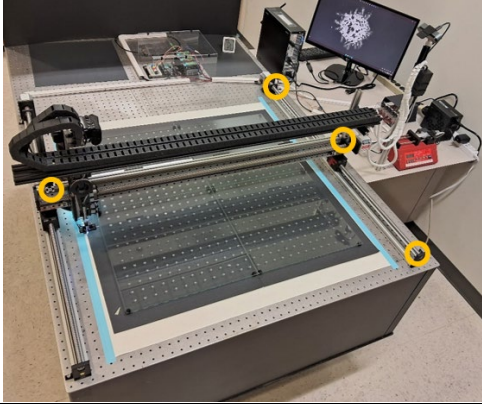
Software Versions Used

MATLAB R2022a / MATLAB R2023a

Python 3.8 (<https://www.python.org/downloads/release/python-380/>). Required Python packages: NumPy and XGBoost.

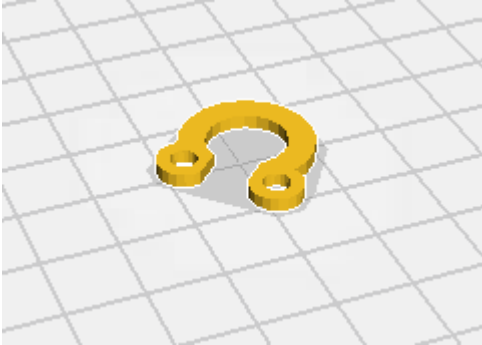
Description of 3D models

01_Stage_Posts	
Model	Location
	
<p>Description: 1 x 1 x 1 cm³ posts used to support and elevate glass plates where the salt solution drops are placed Approximate printing time: 1h45</p>	

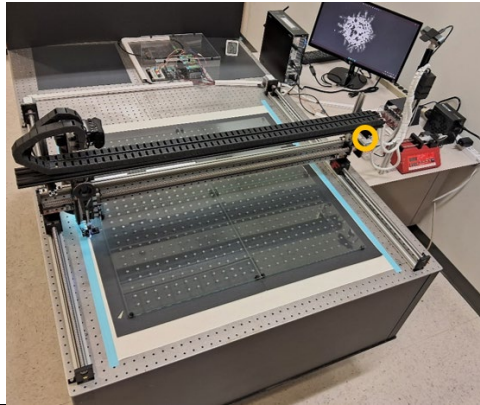
01_Stage_Stopper_Holder	
Model	Location
	
<p>Description: 4 cm piece used to hold stop limit switches at the edges of the linear guides Approximate printing time: 11 min</p>	

02_Belt_Cable_Holder_01

Model



Location

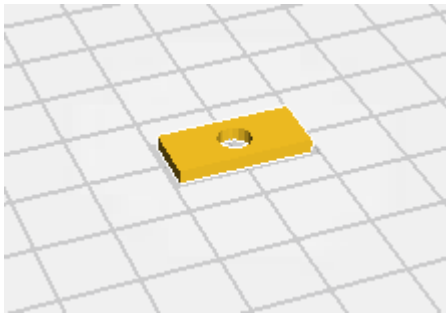


Description: 2 x 2 cm² piece used to secure the electrical cables and tubing to the moving Y linear guide

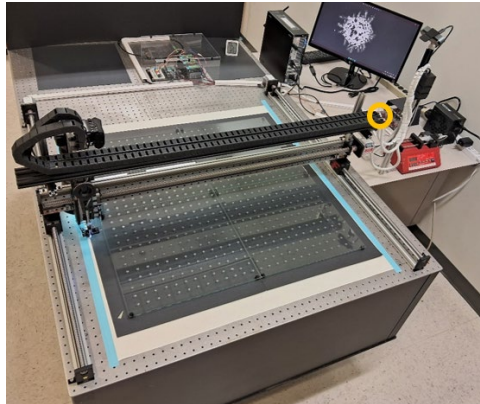
Approximate printing time: 8 min

02_Belt_Cable_Holder_02

Model



Location

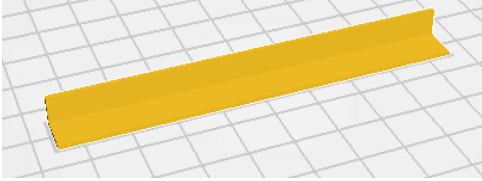


Description: 2 x 1 cm² piece used to secure the electrical cables and tubing to the carrying belt

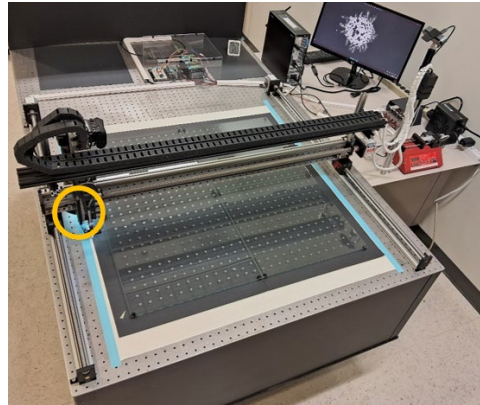
Approximate printing time: 5 min

03_Delivery_Cable_Cover

Model



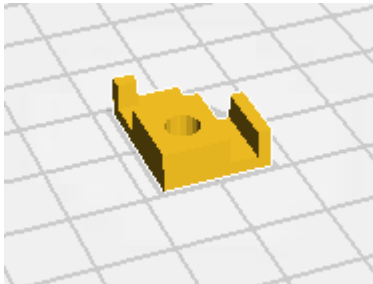
Location



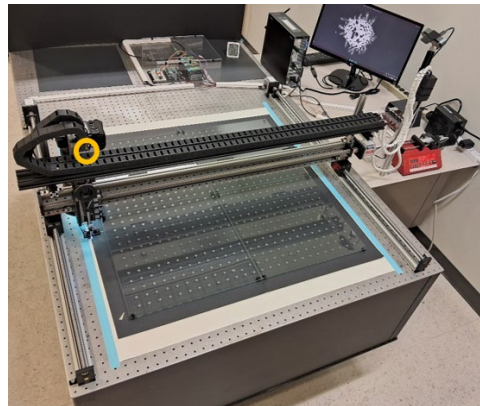
Description: 9 cm cover that protects electrical wiring next to RODI's photosensor
Approximate printing time: 23 min

03_Delivery_Cable_Holder

Model



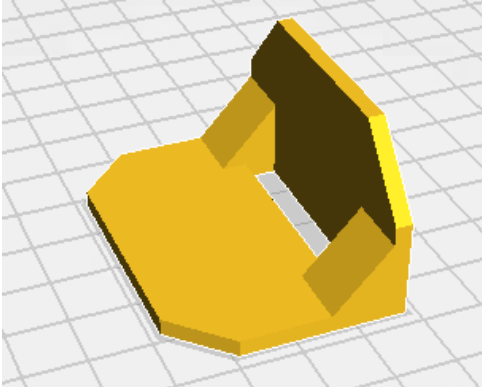
Location



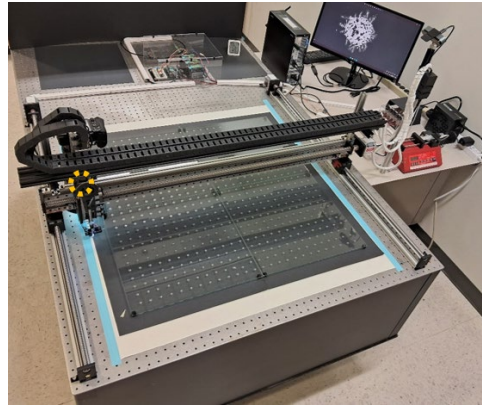
Description: 2 x 2 cm² piece used to secure the camera cables
Approximate printing time: 10 min

03_Delivery_Connector

Model



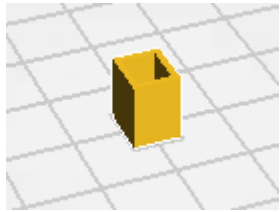
Location



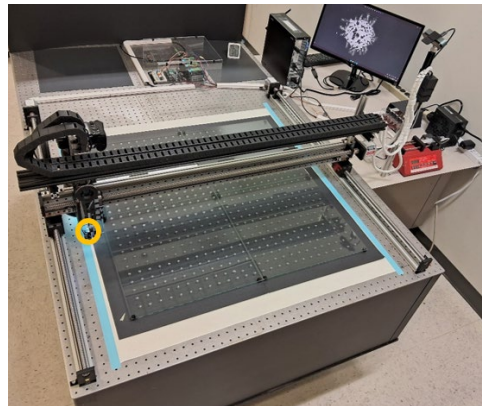
Description: $6 \times 4 \times 4 \text{ cm}^3$ piece that connects the pipette mount to the Y linear guide
Approximate printing time: 1h22

03_Delivery_Photosensor_Encasing

Model



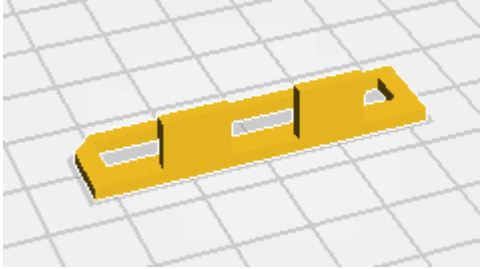
Location



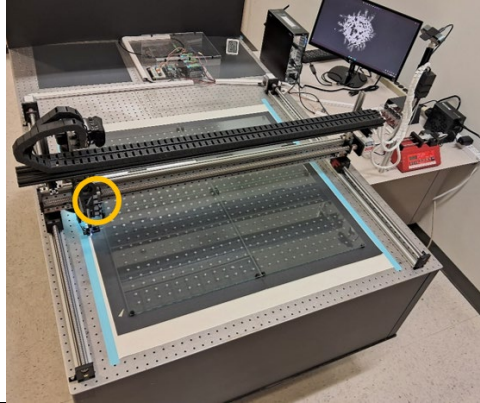
Description: $0.6 \times 0.6 \times 1 \text{ cm}^3$ cap that holds RODI's photosensor or LED
Approximate printing time: 5 min

03_Delivery_Photosensor_Holder_01

Model



Location

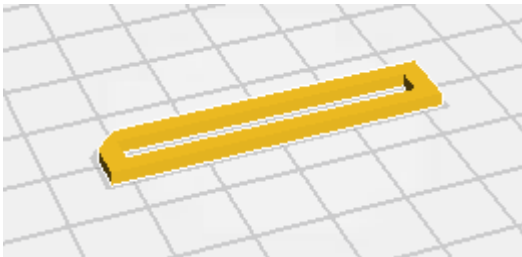


Description: 4.5 x 1.2 cm² piece used to secure the plexiglass mount that holds the pipette

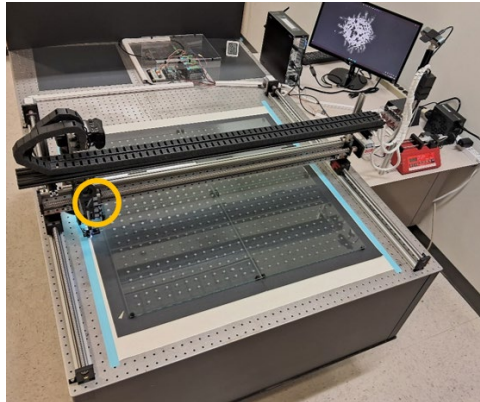
Approximate printing time: 15 min

03_Delivery_Photosensor_Holder_02

Model



Location

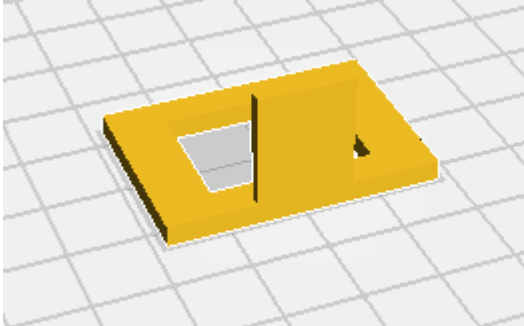


Description: 4.5 x 1.2 cm² piece used to secure the plexiglass mount that holds the pipette

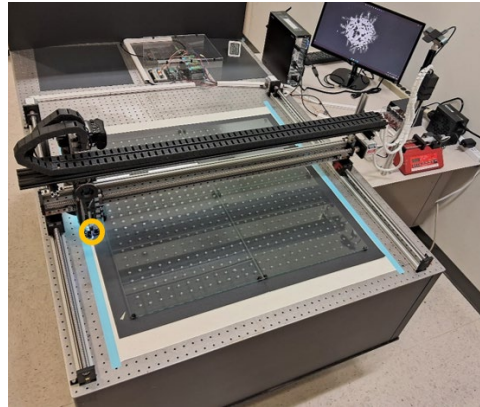
Approximate printing time: 11 min

03_Delivery_Photosensor_Holder_03

Model



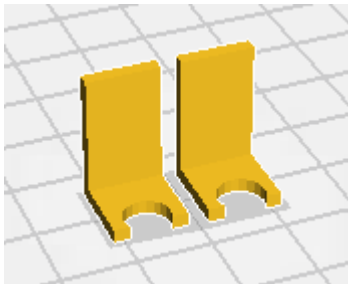
Location



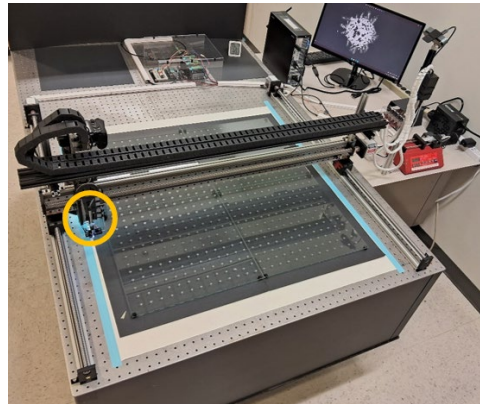
Description: 3.6 x 2.6 cm² mount that supports the photosensing system
Approximate printing time: 26 min

03_Delivery_Pipette_Holder

Model



Location



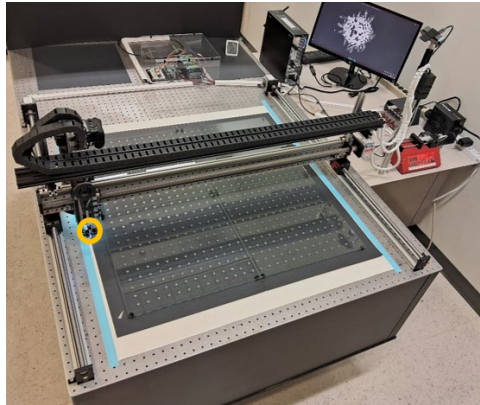
Description: 1 x 1 x 2 cm³ holder that secures the pipette in place
Approximate printing time: 14 min

03_Delivery_PipetteTip_Connector

Model



Location

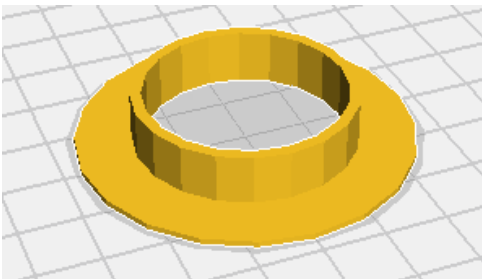


Description: Assortment of pieces that connect the plastic pipette tip to the glass pipette body

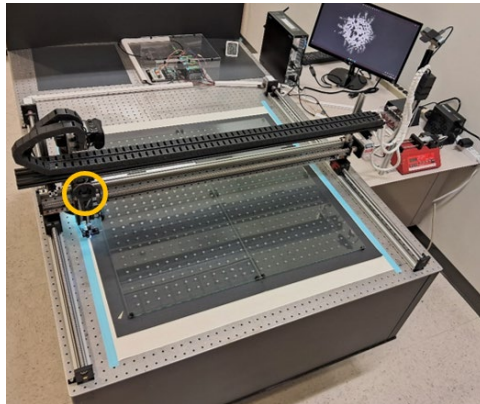
Approximate printing time: 6 min

03_Delivery_Tubing_Carroussel

Model



Location

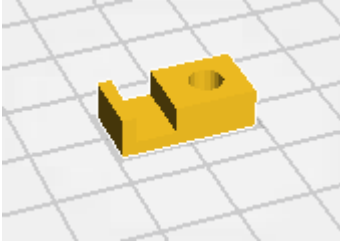


Description: 5 x 5 cm² mount that holds the tubing in place next to the delivery system

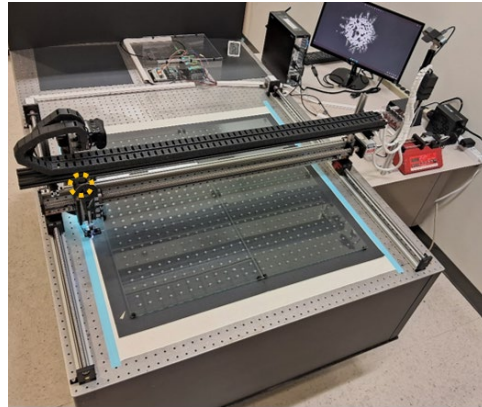
Approximate printing time: 45 min

03_Delivery_Tubing_Holder

Model



Location

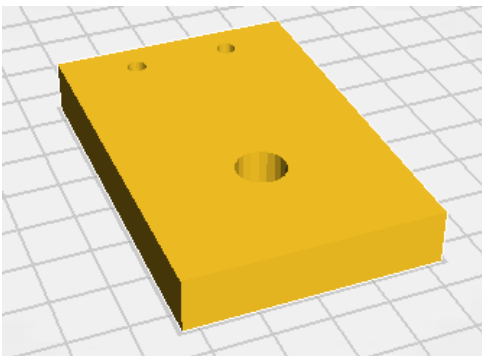


Description: 2 x 1 cm² holder that secure the tubing in place behind the delivery system

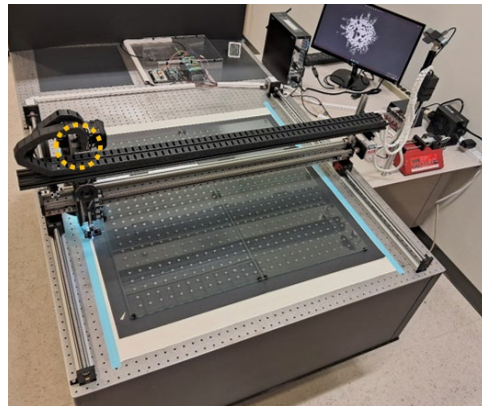
Approximate printing time: 9 min

04_Camera_Connector

Model



Location

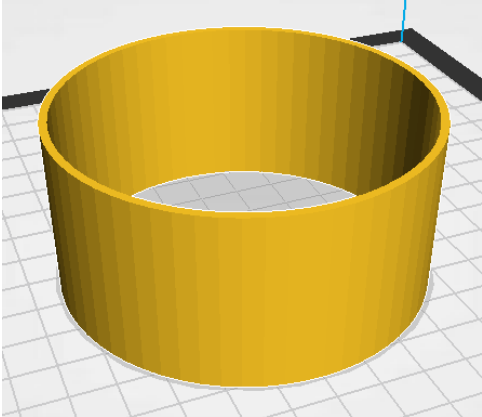


Description: 7 x 4.5 cm² mount that connects the camera to the Y linear guide

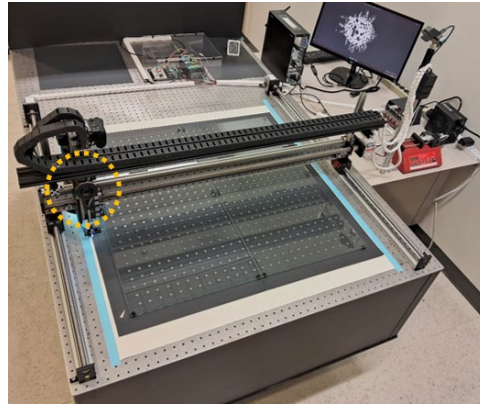
Approximate printing time: 2h46

04_Camera_LED_01

Model



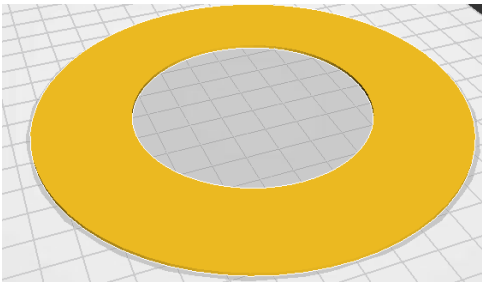
Location



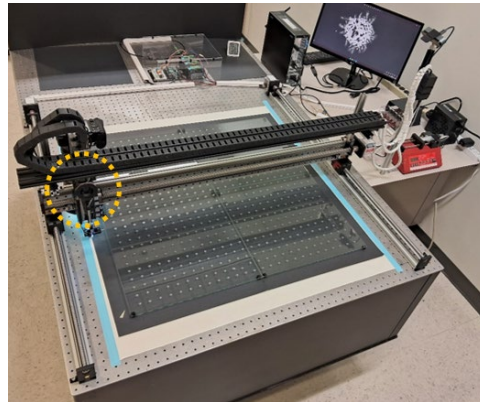
Description: $9 \times 9 \times 5 \text{ cm}^3$ component of the illuminating system
Approximate printing time: 3h36

04_Camera_LED_02

Model



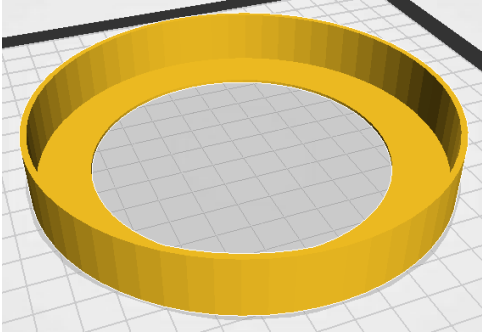
Location



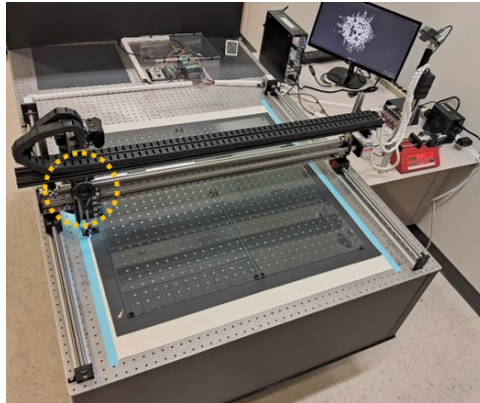
Description: $14.5 \times 14.5 \text{ cm}^2$ component of the illuminating system
Approximate printing time: 2h25

04_Camera_LED_03

Model



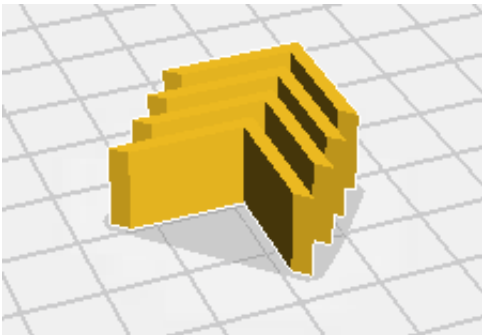
Location



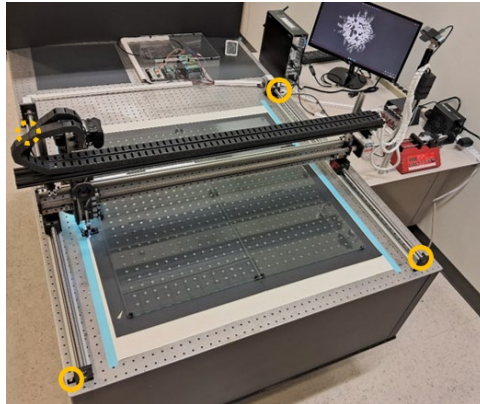
Description: $14 \times 14 \times 2 \text{ cm}^3$ component of the illuminating system
Approximate printing time: 3h42

05_Others_General_Connectors

Model



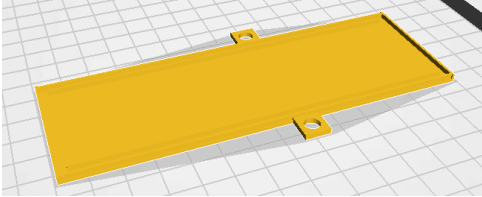
Location



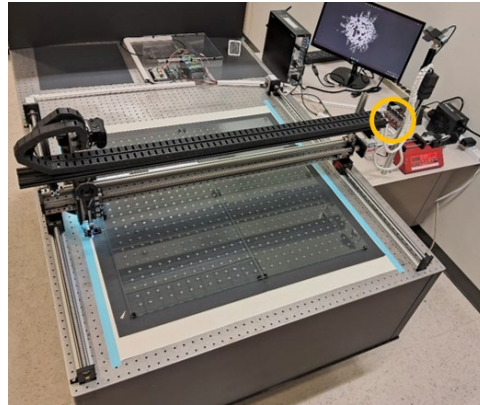
Description: $1.5 \times 1.5 \times 1 \text{ cm}^3$ pieces used to secure the X linear guides to the table
Approximate printing time: 33 min

05_Others_PowerButton_Base

Model



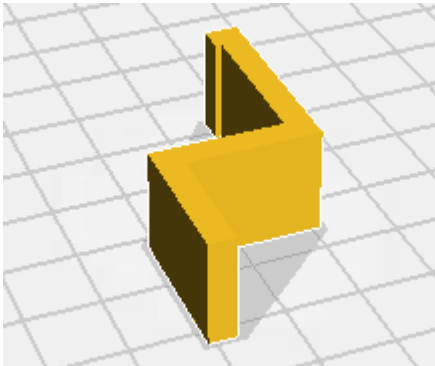
Location



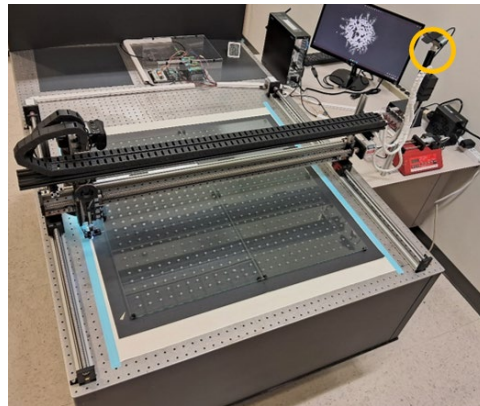
Description: 14 x 5 cm² base that holds the power button box in place
Approximate printing time: 1h57

05_Others_Webcam_Connector

Model



Location



Description: 5 x 2.5 x 1.6 cm³ piece that secures the webcam in place
Approximate printing time: 33 min