

QuPAINT: Physics-Aware Instruction Tuning Approach to Quantum Material Discovery

Supplementary Material

9. Proof of Eq. (3)

Proof (First-order Link to Optical Model). Let $\Delta R(\lambda) = R(x, \lambda) - R_{\text{bg}}(\lambda)$ denote the spectral difference caused by the 2D flake, which depends on the refractive index $n(\lambda)$, thickness d , and interference phase $\phi = 2\pi n(\lambda)d/\lambda$. Assuming $\|\Delta R\|_\infty$ is small (a valid approximation for mono- and few-layer flakes), we can linearize $\Phi \circ \mathbf{I}$ about \mathbf{I}_{bg} using a first-order Taylor expansion:

$$\Delta E(x) \approx \underbrace{\|J_{\Phi}(\mathbf{I}_{\text{bg}}) \int_{\Delta \mathbf{I}} \Delta R(\lambda) E(\lambda) \mathbf{S}(\lambda) d\lambda\|_2}_{\Delta \mathbf{I}}, \quad (5)$$

where J_{Φ} is the Jacobian of the RGB-to-LAB transformation at \mathbf{I}_{bg} and $\mathbf{S}(\lambda) = (S_R(\lambda), S_G(\lambda), S_B(\lambda))^T$. Therefore, there exists a positive constant c (dependent on E, \mathbf{S}, J_{Φ}) such that

$$\Delta E(x) \geq c \|\Delta \mathbf{I}\|_2 = c \left\| \int \Delta R(\lambda) E(\lambda) \mathbf{S}(\lambda) d\lambda \right\|_2. \quad (6)$$

Since the thin-film interference model predicts that $\Delta R(\lambda; d)$ varies systematically with the number of layers d , the perceptual contrast $\Delta E(x)$ is a monotonic proxy (locally) for the reflectance deviation induced by optical interference. Thus, maximizing ΔE over image patches identifies regions whose spectra differ from the substrate due to interference, without explicitly solving the transfer-matrix model.

10. More Details of Synthia

In this section, we further discuss in detail the Synthia framework.

10.1. White Balance Calibration

After the reflected light is captured by the microscope, users often perform a manual color calibration process, typically by applying white balance correction to the image. This step adjusts the gain factors of the red, green, and blue (RGB) channels to compensate for color casts introduced by the illumination source or optical path. In practice, this calibration alters the perceived color of the substrate and flakes. From the perspective of material scientists, such calibration serves two main purposes: (1) it enhances the visibility of the *target flakes*, making them more distinguishable from the background and other overlapping flakes; and (2) each user tends to develop personalized calibration preferences based on their experience and specific material systems, optimizing contrast for their observation workflow. As a result, microscopy images collected across different laboratories often exhibit noticeable color variations, even when

using the same materials, substrates, and hardware configurations. This variability poses a significant challenge for AI models, which must remain robust to inconsistent color calibration settings. Therefore, in our synthetic pipeline, we explicitly model color calibration variability by simulating diverse white balance transformations. This step ensures that the generated synthetic flakes maintain realistic color contrast relative to their substrates and better reflect the diversity of real-world microscopy datasets.

Implementation of White-Balance Calibration. Let the reference microscopy image be $\mathbf{I}_{\text{ref}} \in \mathbb{R}^{3 \times H \times W}$ in linear RGB space, captured for a known material and substrate configuration (e.g., MoS₂ on Si/SiO₂ with known oxide thickness). We denote by $\mathbf{M}_{\text{sub}} \in \{0, 1\}^{H \times W}$ the binary mask identifying substrate regions, and by $\mathbf{1}\{\cdot\}$ the indicator function. The mean substrate color after user calibration (white balance already applied) is computed as:

$$[\mathbf{c}_{\text{sub}}^{\text{ref}}]_k = \frac{\sum_{i,j} \mathbf{I}_{\text{ref}}[k, i, j] \mathbf{M}_{\text{sub}}[i, j]}{\sum_{i,j} \mathbf{M}_{\text{sub}}[i, j]}, \quad k \in \{1, 2, 3\}. \quad (7)$$

Given the known optical stack (air / flake / SiO₂ / Si), the transfer-matrix module \mathcal{T} computes the substrate reflectance spectrum $\mathbf{R}_{\text{sub}}(\lambda) \in \mathbb{R}^D$ sampled across D discrete wavelengths. A colorimetric projection function Φ (CIE 1931 color-matching with standard illuminant) maps spectra to linear RGB, yielding the pre-white-balance substrate color:

$$\mathbf{c}_{\text{sub}}^0 = \Phi(\mathbf{R}_{\text{sub}}(\lambda)) \in \mathbb{R}^3. \quad (8)$$

Assuming a per-channel diagonal white balance model, the user’s gain vector $\mathbf{g} \in \mathbb{R}_+^3$ satisfies:

$$\mathbf{c}_{\text{sub}}^{\text{ref}} \approx \text{diag}(\mathbf{g}) \mathbf{c}_{\text{sub}}^0 \implies \mathbf{g} = \mathbf{c}_{\text{sub}}^{\text{ref}} \oslash \mathbf{c}_{\text{sub}}^0, \quad (9)$$

where \oslash denotes element-wise division (optionally normalized, e.g., $\frac{1}{3} \|\mathbf{g}\|_1 = 1$ to fix global exposure).

Finally, for any synthetic image before white balance $\mathbf{I}_{\text{syn}}^0 \in \mathbb{R}^{3 \times h \times w}$, the personalized correction is applied as:

$$\mathbf{I}_{\text{syn}}[k, i, j] = g_k \cdot \mathbf{I}_{\text{syn}}^0[k, i, j], \quad k \in \{1, 2, 3\}, \quad (10)$$

followed by clipping to $[0, 1]$. This process aligns the synthetic color distribution to the user’s individual calibration preferences while preserving the physically predicted flake–substrate contrast computed by \mathcal{T} and Φ .

10.2. Substrate-Aware Synthetic Placement

Previous data synthesis pipelines often overlook the spatial context of existing flakes in the reference microscopy images. As a result, newly generated synthetic flakes are randomly overlaid on the image, frequently overlapping with real flakes or debris already present on the substrate. Such unawareness introduces unrealistic artifacts and degrades the physical plausibility of the synthesized dataset. A straightforward solution would be to manually annotate bounding boxes for all visible flakes and restrict synthetic placement to the remaining background areas. However, this approach is highly labor-intensive, prone to human error, and does not scale to large microscopy collections. To overcome these limitations, we propose an automated *substrate-detection* algorithm that localizes clean substrate regions directly from the input image. By identifying substrate-only areas through reflectance consistency and color uniformity, our method ensures that synthetic flakes are placed exclusively on physically valid regions, avoiding overlaps with existing structures. This substrate-aware synthesis greatly improves dataset realism and scalability without requiring manual annotations.

10.3. Substrate-Aware Synthetic Flake Generation

Given an unlabeled microscopy image $\mathbf{I} \in [0, 255]^{H \times W \times 3}$, our goal is to synthesize new flakes while avoiding overlap with existing ones. We first compute a LAB-based attention map $\mathbf{A} = \mathcal{F}_{\text{LAB}}(\mathbf{I}) \in \mathbb{R}^{H \times W}$ and normalize it to the range $[0, 1]$:

$$\tilde{\mathbf{A}} = \frac{\mathbf{A} - \min(\mathbf{A})}{\max(\mathbf{A}) - \min(\mathbf{A}) + \varepsilon}. \quad (11)$$

High-attention regions correspond to existing flakes or high-texture areas. The substrate mask is then defined as

$$\mathbf{M}_{\text{sub}} = \mathbf{1} \left\{ \tilde{\mathbf{A}} < \tau \right\}, \quad \tau = \text{Perc}_{90}(\tilde{\mathbf{A}}), \quad (12)$$

where τ denotes the 90th percentile threshold, preserving only low-attention (substrate) regions.

For each synthetic flake f_i with binary mask \mathbf{M}_i and sampled thickness t_i , we randomly select a valid substrate region (u_i, v_i) that satisfies

$$\mathbf{M}_{\text{sub}}[v_i : v_i + h_i, u_i : u_i + w_i] \odot \mathbf{M}_i \equiv 0, \quad (13)$$

ensuring no overlap with existing flakes or previously placed synthetic ones.

Optical Color Generation. The color of each flake is physically computed using the *transfer matrix method* (TMM). Given a material s (e.g., MoS_2), a layer stack $\mathcal{S} = \{\text{air}, s, \text{SiO}_2, \text{Si}\}$, and a flake thickness t_i , the wavelength-dependent reflectance spectrum is obtained by

$$R_i(\lambda) = \mathcal{T}(\mathcal{S}, t_i, \lambda), \quad (14)$$

where \mathcal{T} denotes the multilayer optical simulation. The resulting reflectance is converted into a perceptual RGB color via CIE 1931 color-matching functions $S(\lambda)$ and illuminant spectrum $I(\lambda)$:

$$\mathbf{c}_i = \Phi(R_i(\lambda), I(\lambda)) = \int_{\lambda_{\min}}^{\lambda_{\max}} S(\lambda) I(\lambda) R_i(\lambda) d\lambda, \quad (15)$$

where $\mathbf{c}_i \in [0, 255]^3$ represents the flake’s RGB color.

Flake Composition. The synthetic flake is placed directly on the selected substrate area by replacing the corresponding region in \mathbf{I} :

$$\begin{aligned} \mathbf{I}_{\text{out}}[v_i : v_i + h_i, u_i : u_i + w_i] &= \mathbf{c}_i \mathbf{M}_i \\ &+ \mathbf{I}[v_i : v_i + h_i, u_i : u_i + w_i] (1 - \mathbf{M}_i), \end{aligned} \quad (16)$$

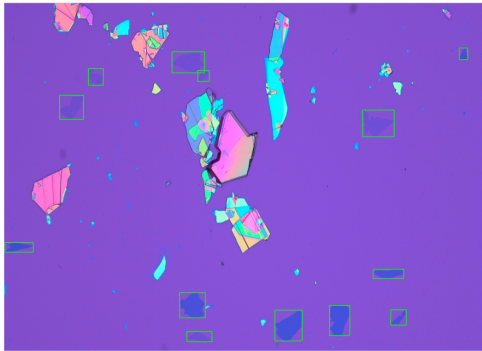
where $\mathbf{M}_i \in \{0, 1\}^{h_i \times w_i}$ denotes the binary flake mask. This substrate-aware process ensures that synthetic flakes are placed only on physically valid substrate regions, avoiding overlaps and preserving realistic color consistency through the optical model.

11. QMat-Instruct Dataset

We present samples from the QMat-Instruct dataset in Fig. 6 through Fig. 15, showing the dialogue between a human and the model concerning material analysis using synthetic optical microscopy images.

12. QF-Bench Dataset

Due to the large size of the data, we currently release 50 samples from the QF-Bench dataset. We show some samples from the benchmark dataset in Fig. 16. The benchmark release is structured in COCO JSON format. The JSON contains top-level keys such as `images`, `annotations`, and `categories`. The dataset defines three distinct categories based on flake thickness: `Mono`, `Few`, and `Thick`, all grouped under the `flake` supercategory. Each individual flake has a bounding box annotation.



Conversation

👤: Does this image contain any mono layer flakes?

🤖: After detecting flakes ([1706, 174, 1738, 219], [629, 189, 751, 270], [316, 254, 372, 318], [726, 261, 769, 303], [208, 357, 298, 450], [1344, 414, 1462, 517], [3, 926, 109, 963], [1383, 1029, 1498, 1065], [658, 1119, 754, 1217], [1219, 1169, 1296, 1285], [1449, 1186, 1507, 1245], [1015, 1187, 1117, 1306], [685, 1270, 779, 1309]) from top-left to bottom-right, I check whether at least one shows very low contrast and semi-transparent properties. If so, mono layers are present; otherwise not. The answer is: Yes.

👤: Count the number of mono layer flakes in this image.

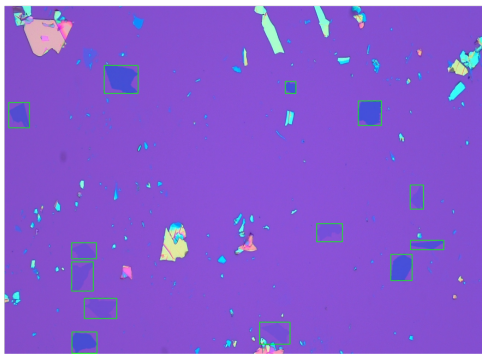
🤖: All flakes are first proposed as [1706, 174, 1738, 219], [629, 189, 751, 270], [316, 254, 372, 318], [726, 261, 769, 303], [208, 357, 298, 450], [1344, 414, 1462, 517], [3, 926, 109, 963], [1383, 1029, 1498, 1065], [658, 1119, 754, 1217], [1219, 1169, 1296, 1285], [1449, 1186, 1507, 1245], [1015, 1187, 1117, 1306], [685, 1270, 779, 1309] from top-left to bottom-right. I then keep only those with faint contrast and low color variation from the substrate to identify mono layers. There are 6 at [629, 189, 751, 270], [316, 254, 372, 318], [726, 261, 769, 303], [208, 357, 298, 450], [1344, 414, 1462, 517], [685, 1270, 779, 1309].

👤: Locate all the mono layer flakes in the image.

🤖: I extract all candidate flakes as [1706, 174, 1738, 219], [629, 189, 751, 270], [316, 254, 372, 318], [726, 261, 769, 303], [208, 357, 298, 450], [1344, 414, 1462, 517], [3, 926, 109, 963], [1383, 1029, 1498, 1065], [658, 1119, 754, 1217], [1219, 1169, 1296, 1285], [1449, 1186, 1507, 1245], [1015, 1187, 1117, 1306], [685, 1270, 779, 1309] from top-left to bottom-right. Mono layers are those with minimal brightness difference and semi-transparent edges. The mono layer flakes are at: [629, 189, 751, 270], [316, 254, 372, 318], [726, 261, 769, 303], [208, 357, 298, 450], [1344, 414, 1462, 517], [685, 1270, 779, 1309]

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 6. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤 : Does this image contain any mono layer flakes?

🤖 : After detecting flakes ([374, 230, 501, 338], [1053, 292, 1093, 338], [1327, 367, 1414, 462], [16, 373, 94, 471], [1522, 692, 1572, 785], [1171, 841, 1268, 911], [1523, 906, 1648, 942], [252, 915, 345, 977], [1449, 960, 1529, 1061], [252, 989, 333, 1102], [299, 1130, 421, 1208], [957, 1222, 1071, 1307], [253, 1260, 347, 1342]) from top-left to bottom-right, I check whether at least one shows very low contrast and semi-transparent properties. If so, mono layers are present; otherwise not. The answer is: Yes.

👤 : How many mono layer flakes are there in the image?

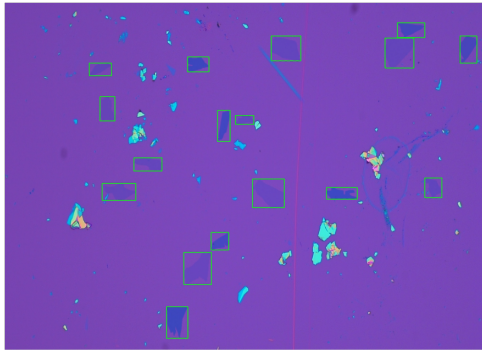
🤖 : I first detect all flake-like regions in the image, producing candidate boxes sorted from top-left to bottom-right [374, 230, 501, 338], [1053, 292, 1093, 338], [1327, 367, 1414, 462], [16, 373, 94, 471], [1522, 692, 1572, 785], [1171, 841, 1268, 911], [1523, 906, 1648, 942], [252, 915, 345, 977], [1449, 960, 1529, 1061], [252, 989, 333, 1102], [299, 1130, 421, 1208], [957, 1222, 1071, 1307], [253, 1260, 347, 1342]. Mono layer flakes usually appear faint, with very low contrast compared to the background and a semi-transparent look. By filtering candidates with these cues, my conclusion is: There are 6 at [1522, 692, 1572, 785], [1171, 841, 1268, 911], [252, 915, 345, 977], [252, 989, 333, 1102], [299, 1130, 421, 1208], [957, 1222, 1071, 1307].

👤 : Where are mono layer flakes?

🤖 : I begin by detecting all flakes in the image from top-left to bottom-right ([374, 230, 501, 338], [1053, 292, 1093, 338], [1327, 367, 1414, 462], [16, 373, 94, 471], [1522, 692, 1572, 785], [1171, 841, 1268, 911], [1523, 906, 1648, 942], [252, 915, 345, 977], [1449, 960, 1529, 1061], [252, 989, 333, 1102], [299, 1130, 421, 1208], [957, 1222, 1071, 1307], [253, 1260, 347, 1342]). Then I select the ones with very low contrast and semi-transparent appearance.. The mono layer flakes are at: [1522, 692, 1572, 785], [1171, 841, 1268, 911], [252, 915, 345, 977], [252, 989, 333, 1102], [299, 1130, 421, 1208], [957, 1222, 1071, 1307]

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 7. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Are there mono layer flakes visible in this sample?

🤖: All flakes are localized as [1469, 78, 1572, 135], [997, 129, 1108, 225], [1704, 129, 1767, 234], [1423, 137, 1531, 253], [684, 210, 764, 268], [316, 234, 399, 283], [356, 364, 412, 458], [796, 417, 843, 537], [863, 437, 932, 471], [483, 601, 588, 652], [1572, 679, 1636, 755], [927, 683, 1046, 794], [364, 700, 490, 766], [1204, 716, 1319, 761], [772, 890, 838, 958], [670, 967, 773, 1092], [605, 1178, 685, 1300] from top-left to bottom-right. By testing their transparency and brightness difference, I confirm whether mono layers are present.. The answer is: Yes.

👤: Would you tell me how many mono layer flakes are there in the image?

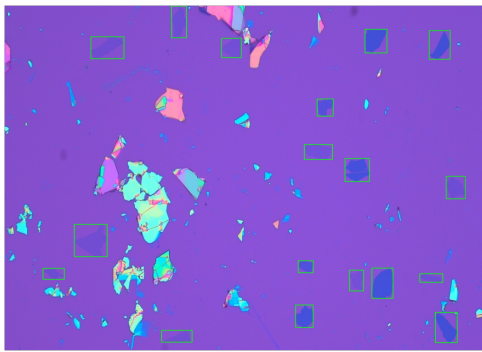
🤖: From the full set of flakes [1469, 78, 1572, 135], [997, 129, 1108, 225], [1704, 129, 1767, 234], [1423, 137, 1531, 253], [684, 210, 764, 268], [316, 234, 399, 283], [356, 364, 412, 458], [796, 417, 843, 537], [863, 437, 932, 471], [483, 601, 588, 652], [1572, 679, 1636, 755], [927, 683, 1046, 794], [364, 700, 490, 766], [1204, 716, 1319, 761], [772, 890, 838, 958], [670, 967, 773, 1092], [605, 1178, 685, 1300] from top-left to bottom-right, I evaluate transparency and contrast against the background to identify mono layers. There are 10 at [997, 129, 1108, 225], [1423, 137, 1531, 253], [316, 234, 399, 283], [356, 364, 412, 458], [863, 437, 932, 471], [483, 601, 588, 652], [1572, 679, 1636, 755], [927, 683, 1046, 794], [364, 700, 490, 766], [670, 967, 773, 1092].

👤: Locate all the mono layer flakes in the image.

🤖: I extract all candidate flakes as [1469, 78, 1572, 135], [997, 129, 1108, 225], [1704, 129, 1767, 234], [1423, 137, 1531, 253], [684, 210, 764, 268], [316, 234, 399, 283], [356, 364, 412, 458], [796, 417, 843, 537], [863, 437, 932, 471], [483, 601, 588, 652], [1572, 679, 1636, 755], [927, 683, 1046, 794], [364, 700, 490, 766], [1204, 716, 1319, 761], [772, 890, 838, 958], [670, 967, 773, 1092], [605, 1178, 685, 1300] from top-left to bottom-right. Mono layers are those with minimal brightness difference and semi-transparent edges. The mono layer flakes are at: [997, 129, 1108, 225], [1423, 137, 1531, 253], [316, 234, 399, 283], [356, 364, 412, 458], [863, 437, 932, 471], [483, 601, 588, 652], [1572, 679, 1636, 755], [927, 683, 1046, 794], [364, 700, 490, 766], [670, 967, 773, 1092]

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 8. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Does this image contain any mono layer flakes?

🤖: After detecting flakes [625, 2, 681, 126], [1349, 93, 1431, 185], [1589, 98, 1668, 210], [323, 122, 447, 208], [812, 128, 886, 203], [1170, 365, 1230, 432], [1122, 543, 1227, 601], [1274, 598, 1366, 688], [1653, 667, 1724, 755], [261, 855, 384, 981], [1099, 997, 1155, 1044], [1375, 1026, 1454, 1144], [144, 1027, 224, 1070], [1291, 1034, 1344, 1115], [1554, 1048, 1640, 1081], [1089, 1170, 1155, 1258], [1612, 1199, 1694, 1317], [588, 1269, 702, 1315]) from top-left to bottom-right, I check whether at least one shows very low contrast and semi-transparent properties. If so, mono layers are present; otherwise not. The answer is: Yes.

👤: Would you tell me how many mono layer flakes are there in the image?

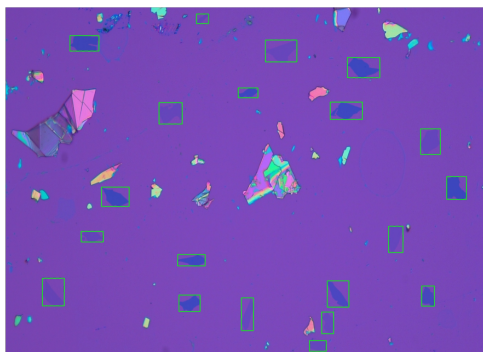
🤖: From the full set of flakes [625, 2, 681, 126], [1349, 93, 1431, 185], [1589, 98, 1668, 210], [323, 122, 447, 208], [812, 128, 886, 203], [1170, 365, 1230, 432], [1122, 543, 1227, 601], [1274, 598, 1366, 688], [1653, 667, 1724, 755], [261, 855, 384, 981], [1099, 997, 1155, 1044], [1375, 1026, 1454, 1144], [144, 1027, 224, 1070], [1291, 1034, 1344, 1115], [1554, 1048, 1640, 1081], [1089, 1170, 1155, 1258], [1612, 1199, 1694, 1317], [588, 1269, 702, 1315] from top-left to bottom-right, I evaluate transparency and contrast against the background to identify mono layers.. There are 10 at [625, 2, 681, 126], [323, 122, 447, 208], [812, 128, 886, 203], [1122, 543, 1227, 601], [1653, 667, 1724, 755], [261, 855, 384, 981], [144, 1027, 224, 1070], [1291, 1034, 1344, 1115], [1554, 1048, 1640, 1081], [588, 1269, 702, 1315].

👤: Locate all the mono layer flakes in the image.

🤖: I extract all candidate flakes as [625, 2, 681, 126], [1349, 93, 1431, 185], [1589, 98, 1668, 210], [323, 122, 447, 208], [812, 128, 886, 203], [1170, 365, 1230, 432], [1122, 543, 1227, 601], [1274, 598, 1366, 688], [1653, 667, 1724, 755], [261, 855, 384, 981], [1099, 997, 1155, 1044], [1375, 1026, 1454, 1144], [144, 1027, 224, 1070], [1291, 1034, 1344, 1115], [1554, 1048, 1640, 1081], [1089, 1170, 1155, 1258], [1612, 1199, 1694, 1317], [588, 1269, 702, 1315] from top-left to bottom-right. Mono layers are those with minimal brightness difference and semi-transparent edges. The mono layer flakes are at: [625, 2, 681, 126], [323, 122, 447, 208], [812, 128, 886, 203], [1122, 543, 1227, 601], [1653, 667, 1724, 755], [261, 855, 384, 981], [144, 1027, 224, 1070], [1291, 1034, 1344, 1115], [1554, 1048, 1640, 1081], [588, 1269, 702, 1315]

Flake Candidate Enumeration Physics-Informed Reasoning Structured Final Conclusion

Figure 9. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Confirm if this sample includes mono layer flakes.

🤖: I detect all flakes ([715, 26, 761, 63], [241, 110, 350, 170], [974, 127, 1092, 212], [1281, 195, 1403, 273], [872, 313, 946, 351], [1217, 368, 1338, 435], [575, 370, 662, 453], [1556, 472, 1630, 571], [1654, 657, 1728, 746], [359, 698, 463, 773], [1436, 850, 1490, 952], [283, 870, 366, 911], [645, 960, 748, 1004], [138, 1052, 220, 1159], [1206, 1062, 1287, 1164], [1560, 1082, 1608, 1160], [650, 1117, 729, 1181], [884, 1128, 929, 1255], [1185, 1183, 1230, 1267], [1139, 1294, 1202, 1335]) and apply mono layer criteria: low contrast and subtle color shift relative to the background. The answer is: Yes.

👤: Count the number of mono layer flakes in this image.

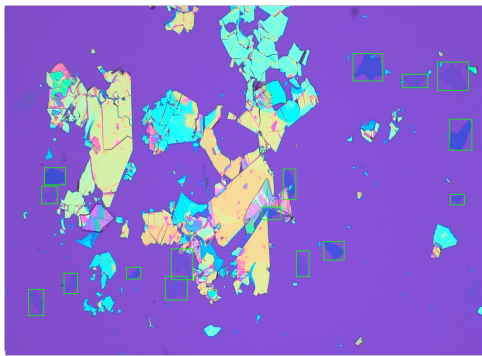
🤖: All flakes are first proposed as [715, 26, 761, 63], [241, 110, 350, 170], [974, 127, 1092, 212], [1281, 195, 1403, 273], [872, 313, 946, 351], [1217, 368, 1338, 435], [575, 370, 662, 453], [1556, 472, 1630, 571], [1654, 657, 1728, 746], [359, 698, 463, 773], [1436, 850, 1490, 952], [283, 870, 366, 911], [645, 960, 748, 1004], [138, 1052, 220, 1159], [1206, 1062, 1287, 1164], [1560, 1082, 1608, 1160], [650, 1117, 729, 1181], [884, 1128, 929, 1255], [1185, 1183, 1230, 1267], [1139, 1294, 1202, 1335] from top-left to bottom-right. I then keep only those with faint contrast and low color variation from the substrate to identify mono layers. There are 10 at [715, 26, 761, 63], [974, 127, 1092, 212], [575, 370, 662, 453], [1556, 472, 1630, 571], [1436, 850, 1490, 952], [283, 870, 366, 911], [138, 1052, 220, 1159], [884, 1128, 929, 1255], [1185, 1183, 1230, 1267], [1139, 1294, 1202, 1335].

👤: Where are mono layer flakes?

🤖: I begin by detecting all flakes in the image from top-left to bottom-right ([715, 26, 761, 63], [241, 110, 350, 170], [974, 127, 1092, 212], [1281, 195, 1403, 273], [872, 313, 946, 351], [1217, 368, 1338, 435], [575, 370, 662, 453], [1556, 472, 1630, 571], [1654, 657, 1728, 746], [359, 698, 463, 773], [1436, 850, 1490, 952], [283, 870, 366, 911], [645, 960, 748, 1004], [138, 1052, 220, 1159], [1206, 1062, 1287, 1164], [1560, 1082, 1608, 1160], [650, 1117, 729, 1181], [884, 1128, 929, 1255], [1185, 1183, 1230, 1267], [1139, 1294, 1202, 1335]). 449, 1186, 1507, 1245], [1015, 1187, 1117, 1306], [685, 1270, 779, 1309] from top-left to bottom-right. Then I select the ones with very low contrast and semi-transparent appearance. The mono layer flakes are at: [715, 26, 761, 63], [974, 127, 1092, 212], [575, 370, 662, 453], [1556, 472, 1630, 571], [1436, 850, 1490, 952], [283, 870, 366, 911], [138, 1052, 220, 1159], [884, 1128, 929, 1255], [1185, 1183, 1230, 1267], [1139, 1294, 1202, 1335].

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 10. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Confirm if this sample includes mono layer flakes.

🤖: I detect all flakes ([[1303, 184, 1415, 290], [1620, 217, 1736, 326], [1488, 264, 1584, 315], [1664, 437, 1749, 556], [147, 623, 226, 689], [1042, 628, 1089, 746], [136, 694, 195, 762], [1669, 726, 1721, 766], [962, 774, 1033, 823], [1194, 907, 1270, 978], [622, 935, 702, 1054], [1091, 943, 1139, 1042], [454, 1006, 508, 1051], [220, 1029, 271, 1107], [600, 1041, 682, 1132], [87, 1091, 144, 1192]]) and apply monolayer criteria: low contrast and subtle color shift relative to the background. The answer is: Yes.

👤: How many mono layer flakes are there in the image?

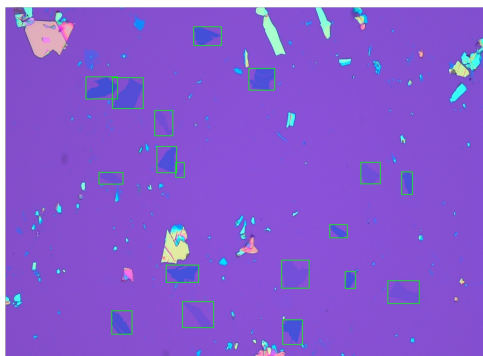
🤖: I first detect all flake-like regions in the image, producing candidate boxes sorted from top-left to bottom-right [1303, 184, 1415, 290], [1620, 217, 1736, 326], [1488, 264, 1584, 315], [1664, 437, 1749, 556], [147, 623, 226, 689], [1042, 628, 1089, 746], [136, 694, 195, 762], [1669, 726, 1721, 766], [962, 774, 1033, 823], [1194, 907, 1270, 978], [622, 935, 702, 1054], [1091, 943, 1139, 1042], [454, 1006, 508, 1051], [220, 1029, 271, 1107], [600, 1041, 682, 1132], [87, 1091, 144, 1192] Monolayer flakes usually appear faint, with very low contrast compared to the background and a semi-transparent look. By filtering candidates with these cues, my conclusion is: There are 8 at [1620, 217, 1736, 326], [1488, 264, 1584, 315], [136, 694, 195, 762], [622, 935, 702, 1054], [1091, 943, 1139, 1042], [220, 1029, 271, 1107], [600, 1041, 682, 1132], [87, 1091, 144, 1192].

👤: Locate all the mono layer flakes in the image.

🤖: I extract all candidate flakes as [1303, 184, 1415, 290], [1620, 217, 1736, 326], [1488, 264, 1584, 315], [1664, 437, 1749, 556], [147, 623, 226, 689], [1042, 628, 1089, 746], [136, 694, 195, 762], [1669, 726, 1721, 766], [962, 774, 1033, 823], [1194, 907, 1270, 978], [622, 935, 702, 1054], [1091, 943, 1139, 1042], [454, 1006, 508, 1051], [220, 1029, 271, 1107], [600, 1041, 682, 1132], [87, 1091, 144, 1192] from top-left to bottom-right. Mono layers are those with minimal brightness difference and semi-transparent edges. The mono layer flakes are at: [1620, 217, 1736, 326], [1488, 264, 1584, 315], [136, 694, 195, 762], [622, 935, 702, 1054], [1091, 943, 1139, 1042], [220, 1029, 271, 1107], [600, 1041, 682, 1132], [87, 1091, 144, 1192]

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 11. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Are there mono layer flakes visible in this sample?

🤖: All flakes are localized as [705, 75, 809, 147], [912, 235, 1009, 319], [301, 268, 420, 353], [401, 271, 516, 390], [559, 397, 627, 495], [567, 536, 641, 637], [1332, 597, 1405, 680], [638, 598, 670, 655], [1485, 634, 1525, 721], [350, 636, 440, 681], [1215, 837, 1281, 888], [1035, 974, 1138, 1082], [602, 993, 724, 1060], [1273, 1018, 1312, 1083], [1433, 1053, 1551, 1141], [663, 1134, 780, 1234], [398, 1169, 474, 1259], [1039, 1204, 1113, 1299] from top-left to bottom-right. By testing their transparency and brightness difference, I confirm whether mono layers are present. The answer is: Yes.

👤: How many mono layer flakes are there in the image?

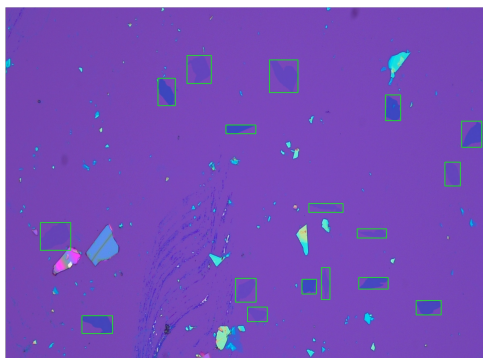
🤖: I first detect all flake-like regions in the image, producing candidate boxes sorted from top-left to bottom-right [705, 75, 809, 147], [912, 235, 1009, 319], [301, 268, 420, 353], [401, 271, 516, 390], [559, 397, 627, 495], [567, 536, 641, 637], [1332, 597, 1405, 680], [638, 598, 670, 655], [1485, 634, 1525, 721], [350, 636, 440, 681], [1215, 837, 1281, 888], [1035, 974, 1138, 1082], [602, 993, 724, 1060], [1273, 1018, 1312, 1083], [1433, 1053, 1551, 1141], [663, 1134, 780, 1234], [398, 1169, 474, 1259], [1039, 1204, 1113, 1299]. Monolayer flakes usually appear faint, with very low contrast compared to the background and a semi-transparent look. By filtering candidates with these cues, my conclusion is: There are 7 at [559, 397, 627, 495], [1332, 597, 1405, 680], [638, 598, 670, 655], [350, 636, 440, 681], [1035, 974, 1138, 1082], [1433, 1053, 1551, 1141], [663, 1134, 780, 1234].

👤: Locate all the mono layer flakes in the image

🤖: I extract all candidate flakes as [705, 75, 809, 147], [912, 235, 1009, 319], [301, 268, 420, 353], [401, 271, 516, 390], [559, 397, 627, 495], [567, 536, 641, 637], [1332, 597, 1405, 680], [638, 598, 670, 655], [1485, 634, 1525, 721], [350, 636, 440, 681], [1215, 837, 1281, 888], [1035, 974, 1138, 1082], [602, 993, 724, 1060], [1273, 1018, 1312, 1083], [1433, 1053, 1551, 1141], [663, 1134, 780, 1234], [398, 1169, 474, 1259], [1039, 1204, 1113, 1299] from top-left to bottom-right. Mono layers are those with minimal brightness difference and semi-transparent edges. The mono layer flakes are at: [559, 397, 627, 495], [1332, 597, 1405, 680], [638, 598, 670, 655], [350, 636, 440, 681], [1035, 974, 1138, 1082], [1433, 1053, 1551, 1141], [663, 1134, 780, 1234]

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 12. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Are there mono layer flakes visible in this sample?

🤖: All flakes are localized as [680, 184, 771, 291], [989, 201, 1096, 326], [571, 271, 636, 376], [1424, 335, 1481, 433], [1710, 435, 1785, 535], [826, 450, 938, 483], [1647, 592, 1705, 683], [1137, 750, 1265, 784], [130, 823, 244, 929], [1318, 847, 1428, 882], [1184, 995, 1216, 1117], [1324, 1033, 1434, 1077], [862, 1036, 939, 1127], [1111, 1040, 1165, 1096], [1539, 1121, 1633, 1176], [907, 1146, 981, 1202], [286, 1179, 401, 1247] from top-left to bottom-right. By testing their transparency and brightness difference, I confirm whether mono layers are present. The answer is: Yes.

👤: Count the number of mono layer flakes in this image.

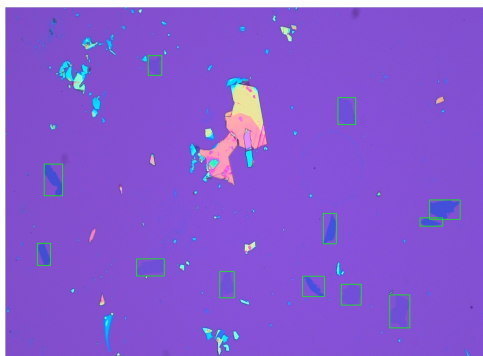
🤖: All flakes are first proposed as [680, 184, 771, 291], [989, 201, 1096, 326], [571, 271, 636, 376], [1424, 335, 1481, 433], [1710, 435, 1785, 535], [826, 450, 938, 483], [1647, 592, 1705, 683], [1137, 750, 1265, 784], [130, 823, 244, 929], [1318, 847, 1428, 882], [1184, 995, 1216, 1117], [1324, 1033, 1434, 1077], [862, 1036, 939, 1127], [1111, 1040, 1165, 1096], [1539, 1121, 1633, 1176], [907, 1146, 981, 1202], [286, 1179, 401, 1247] from top-left to bottom-right. I then keep only those with faint contrast and low color variation from the substrate to identify mono layers. The 9 at [680, 184, 771, 291], [989, 201, 1096, 326], [1647, 592, 1705, 683], [1137, 750, 1265, 784], [130, 823, 244, 929], [1318, 847, 1428, 882], [1184, 995, 1216, 1117], [862, 1036, 939, 1127], [907, 1146, 981, 1202].

👤: Find the bounding box, locations of mono layer flakes.

🤖: All flake regions are extracted as [680, 184, 771, 291], [989, 201, 1096, 326], [571, 271, 636, 376], [1424, 335, 1481, 433], [1710, 435, 1785, 535], [826, 450, 938, 483], [1647, 592, 1705, 683], [1137, 750, 1265, 784], [130, 823, 244, 929], [1318, 847, 1428, 882], [1184, 995, 1216, 1117], [1324, 1033, 1434, 1077], [862, 1036, 939, 1127], [1111, 1040, 1165, 1096], [1539, 1121, 1633, 1176], [907, 1146, 981, 1202], [286, 1179, 401, 1247]. I then check each region: mono layer flakes show lighter contrast and subtle color shift relative to the substrate. [680, 184, 771, 291], [989, 201, 1096, 326], [1647, 592, 1705, 683], [1137, 750, 1265, 784], [130, 823, 244, 929], [1318, 847, 1428, 882], [1184, 995, 1216, 1117], [862, 1036, 939, 1127], [907, 1146, 981, 1202]

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 13. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Are there mono layer flakes visible in this sample?

🤖: All flakes are localized as [533, 186, 584, 262], [1243, 348, 1309, 452], [144, 604, 213, 726], [1586, 742, 1700, 817], [1188, 794, 1237, 911], [1550, 811, 1635, 843], [119, 909, 168, 994], [490, 969, 593, 1035], [801, 1018, 855, 1119], [1111, 1036, 1192, 1114], [1255, 1067, 1330, 1147], [1437, 1108, 1512, 1235] from top-left to bottom-right. . By testing their transparency and brightness difference, I confirm whether mono layers are present. The answer is: Yes.

👤: Would you tell me how many mono layer flakes are there in the image?

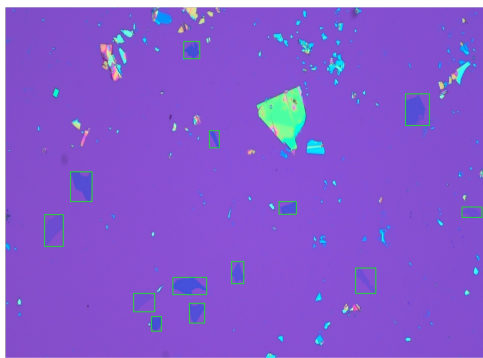
🤖: From the full set of flakes [533, 186, 584, 262], [1243, 348, 1309, 452], [144, 604, 213, 726], [1586, 742, 1700, 817], [1188, 794, 1237, 911], [1550, 811, 1635, 843], [119, 909, 168, 994], [490, 969, 593, 1035], [801, 1018, 855, 1119], [1111, 1036, 1192, 1114], [1255, 1067, 1330, 1147], [1437, 1108, 1512, 1235] from top-left to bottom-right, I evaluate transparency and contrast against the background to identify mono layers. There are 6 at [533, 186, 584, 262], [1243, 348, 1309, 452], [490, 969, 593, 1035], [801, 1018, 855, 1119], [1255, 1067, 1330, 1147], [1437, 1108, 1512, 1235].

👤: Where are mono layer flakes?

🤖: I begin by detecting all flakes in the image from top-left to bottom-right ([533, 186, 584, 262], [1243, 348, 1309, 452], [144, 604, 213, 726], [1586, 742, 1700, 817], [1188, 794, 1237, 911], [1550, 811, 1635, 843], [119, 909, 168, 994], [490, 969, 593, 1035], [801, 1018, 855, 1119], [1111, 1036, 1192, 1114], [1255, 1067, 1330, 1147], [1437, 1108, 1512, 1235]). Then I select the ones with very low contrast and semi-transparent appearance. The mono layer flakes are at: [533, 186, 584, 262], [1243, 348, 1309, 452], [490, 969, 593, 1035], [801, 1018, 855, 1119], [1255, 1067, 1330, 1147], [1437, 1108, 1512, 1235]

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 14. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).



Conversation

👤: Confirm if this sample includes mono layer flakes.

🤖: I detect all flakes ([668, 131, 728, 197], [1499, 332, 1589, 453], [764, 473, 801, 539], [243, 631, 324, 745], [1025, 745, 1090, 794], [1711, 766, 1786, 806], [146, 795, 216, 917], [847, 975, 893, 1060], [1312, 1000, 1388, 1097], [627, 1036, 754, 1102], [480, 1097, 559, 1168], [689, 1135, 746, 1211], [547, 1186, 584, 1242]) and apply mono layer criteria: low contrast and subtle color shift relative to the background. The answer is: Yes.

👤: How many mono layer flakes are there in the image?

🤖: I first detect all flake-like regions in the image, producing candidate boxes sorted from top-left to bottom-right [668, 131, 728, 197], [1499, 332, 1589, 453], [764, 473, 801, 539], [243, 631, 324, 745], [1025, 745, 1090, 794], [1711, 766, 1786, 806], [146, 795, 216, 917], [847, 975, 893, 1060], [1312, 1000, 1388, 1097], [627, 1036, 754, 1102], [480, 1097, 559, 1168], [689, 1135, 746, 1211], [547, 1186, 584, 1242]. Monolayer flakes usually appear faint, with very low contrast compared to the background and a semi-transparent look. By filtering candidates with these cues, my conclusion is: There are 4 at [1711, 766, 1786, 806], [146, 795, 216, 917], [1312, 1000, 1388, 1097], [480, 1097, 559, 1168].

👤: Identify the bounding boxes of the mono layer flakes.

🤖: Flake candidates are first collected ([668, 131, 728, 197], [1499, 332, 1589, 453], [764, 473, 801, 539], [243, 631, 324, 745], [1025, 745, 1090, 794], [1711, 766, 1786, 806], [146, 795, 216, 917], [847, 975, 893, 1060], [1312, 1000, 1388, 1097], [627, 1036, 754, 1102], [480, 1097, 559, 1168], [689, 1135, 746, 1211], [547, 1186, 584, 1242]) from top-left to bottom-right. Mono layers appear as light, near-transparent regions. The mono layer flakes are at: [1711, 766, 1786, 806], [146, 795, 216, 917], [1312, 1000, 1388, 1097], [480, 1097, 559, 1168].

Flake Candidate Enumeration
 Physics-Informed Reasoning
 Structured Final Conclusion

Figure 15. Conversation between a human and model analyzing a synthetic optical microscopy image. The model output is highlighted in three parts: flake candidate enumeration (pink), physics-informed reasoning (blue), and final conclusion (green).

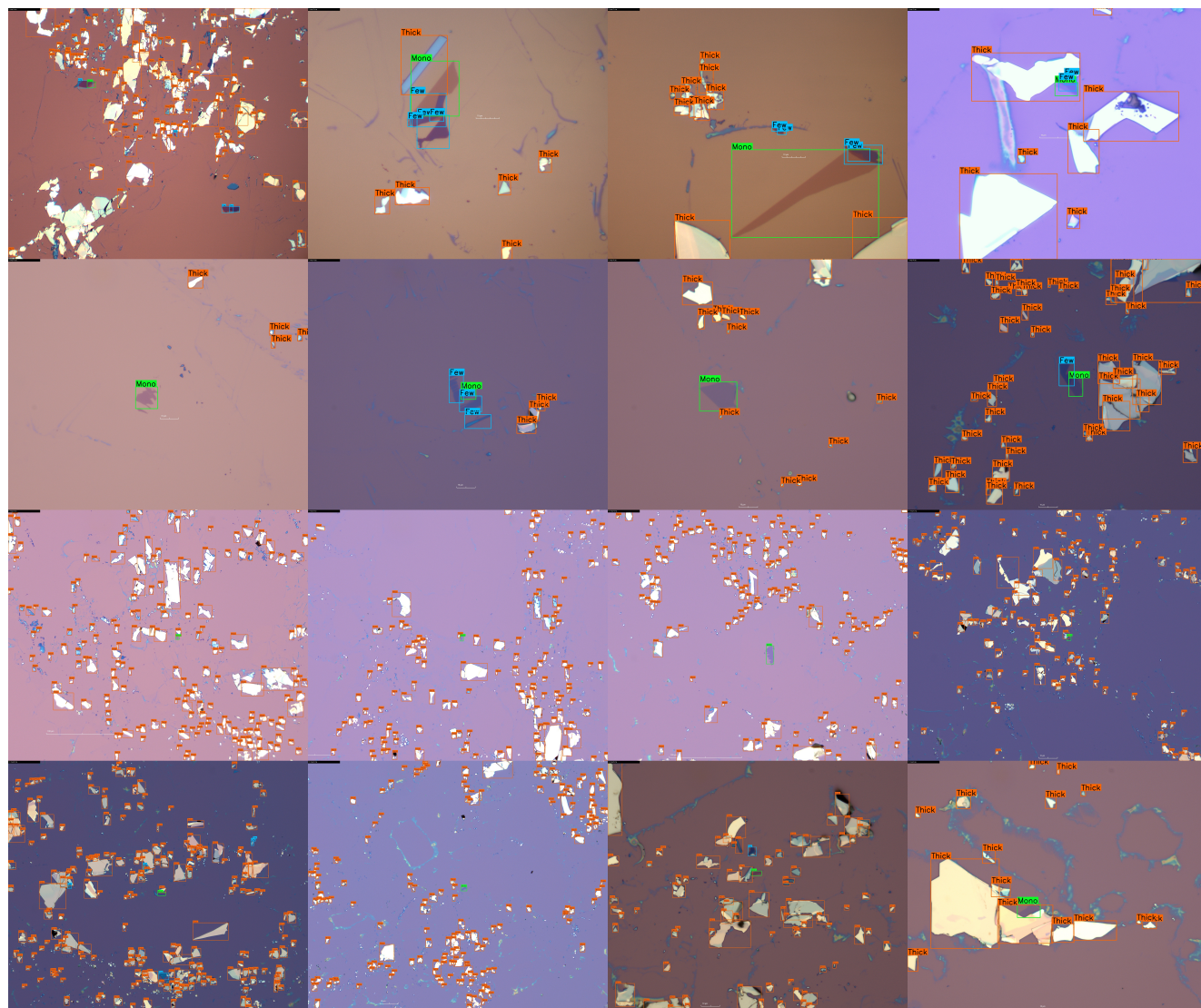


Figure 16. Samples from the QF-Bench dataset. Each microscopy image displays annotations for 2D material flakes. Flake thickness categories are color-coded: green (Mono), blue (Few), and orange (Thick).