Supplementary Information

Identifying Hidden High-Dimensional Structure/Property Relationships Using Self-organising Maps

Amanda S. Barnard,* Benyamin Motevalli, and Baichuan Sun

CSIRO Data61, Docklands, Victoria 3008, Australia

E-mail: amanda.barnard@data61.csiro.au

Contained in this document are histograms of the structural features (Figure S1) and target property labels (Figure S3) of the single crystal and twinned diamond nanostructures defined in the main text. As mentioned in the main text, all of these structures are freely available online. This is followed by additional self-organising maps encoded by the actual magnitude of variables before normalization, both the remaining structural features not shown in the main text (Figure S4), and the remaining target property labels not shown in the main text (Figure S5). The external categorical features, the nanoparticles shapes, are depicted in Figure S6.

*To whom correspondence should be addressed

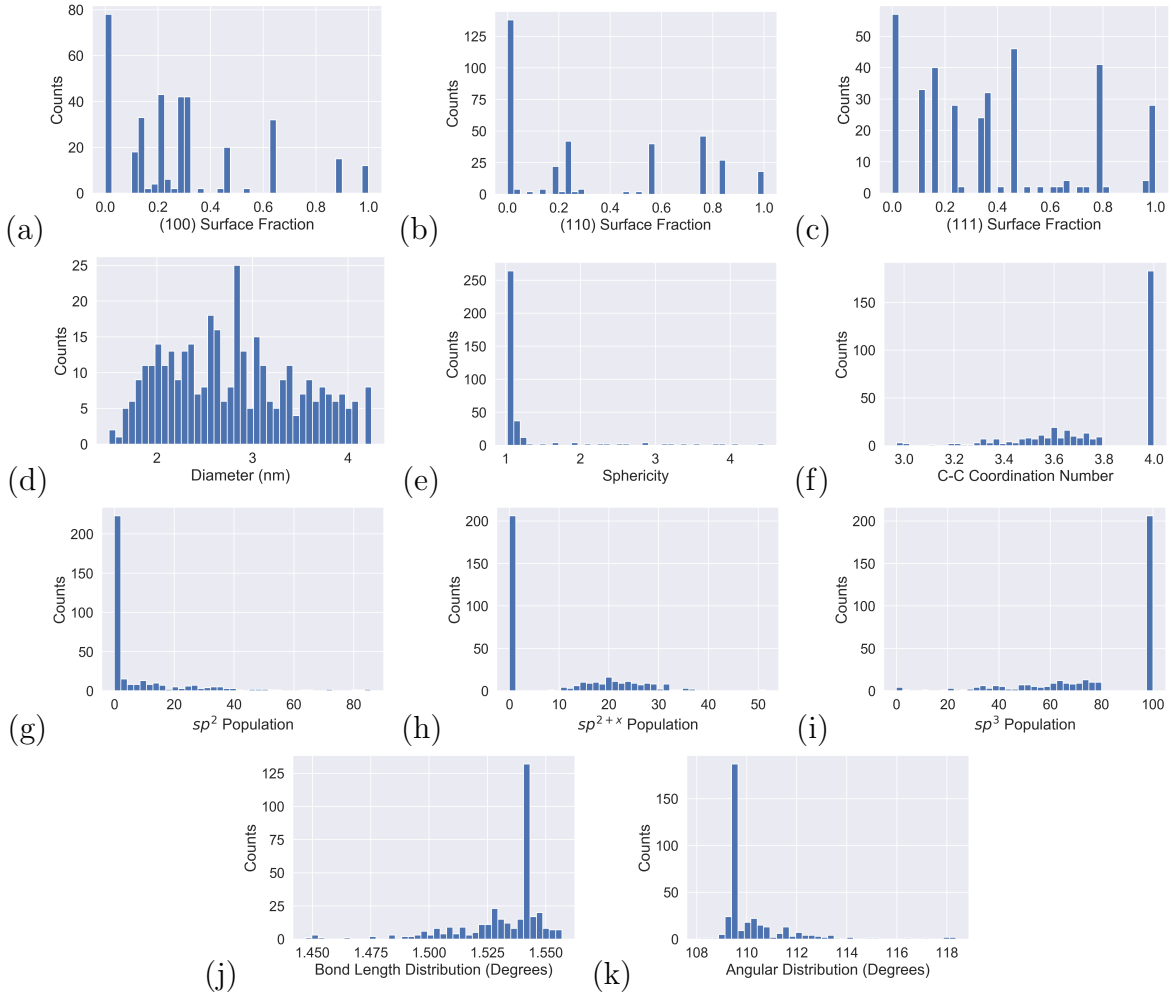

Figure S1: Distribution of structural features within the 355 particle diamond nanoparticle dataset, including the (a) fraction of {100} facet area, (b) fraction of {110} facet area, (c) fraction of {111} facet area, (d) average diameter, (e) sphericity, (f) average C–C coordination number (4 being perfect diamond, and 3 being perfect graphene), (g) fraction of sp^2 hybridized atoms, (h) fraction of sp^{2+x} hybridized atoms, (i) fraction of sp^3 hybridized atoms, (j) average C–C bond length, and (k) the average C–C–C bond angle.

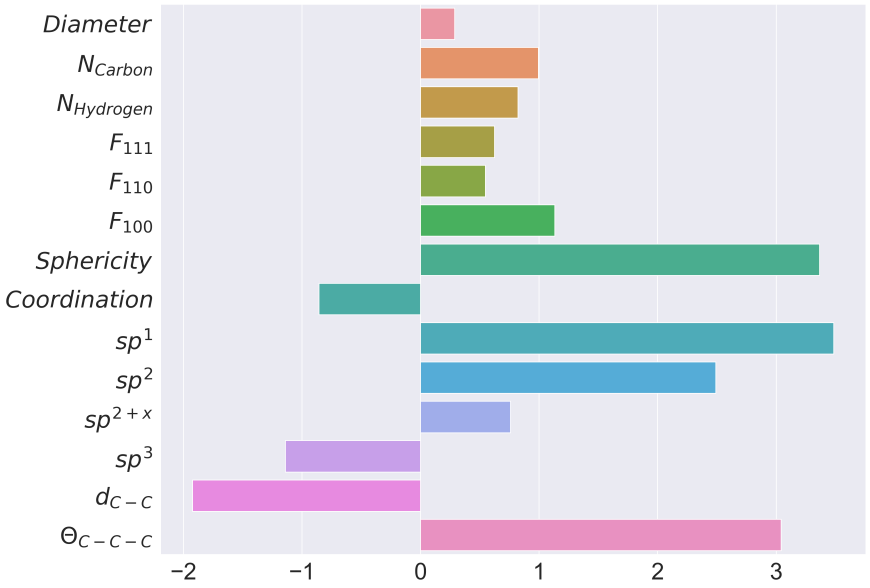

Figure S2: Skew map of the 10 features used to predict the charge transfer properties of the diamond nanoparticles.

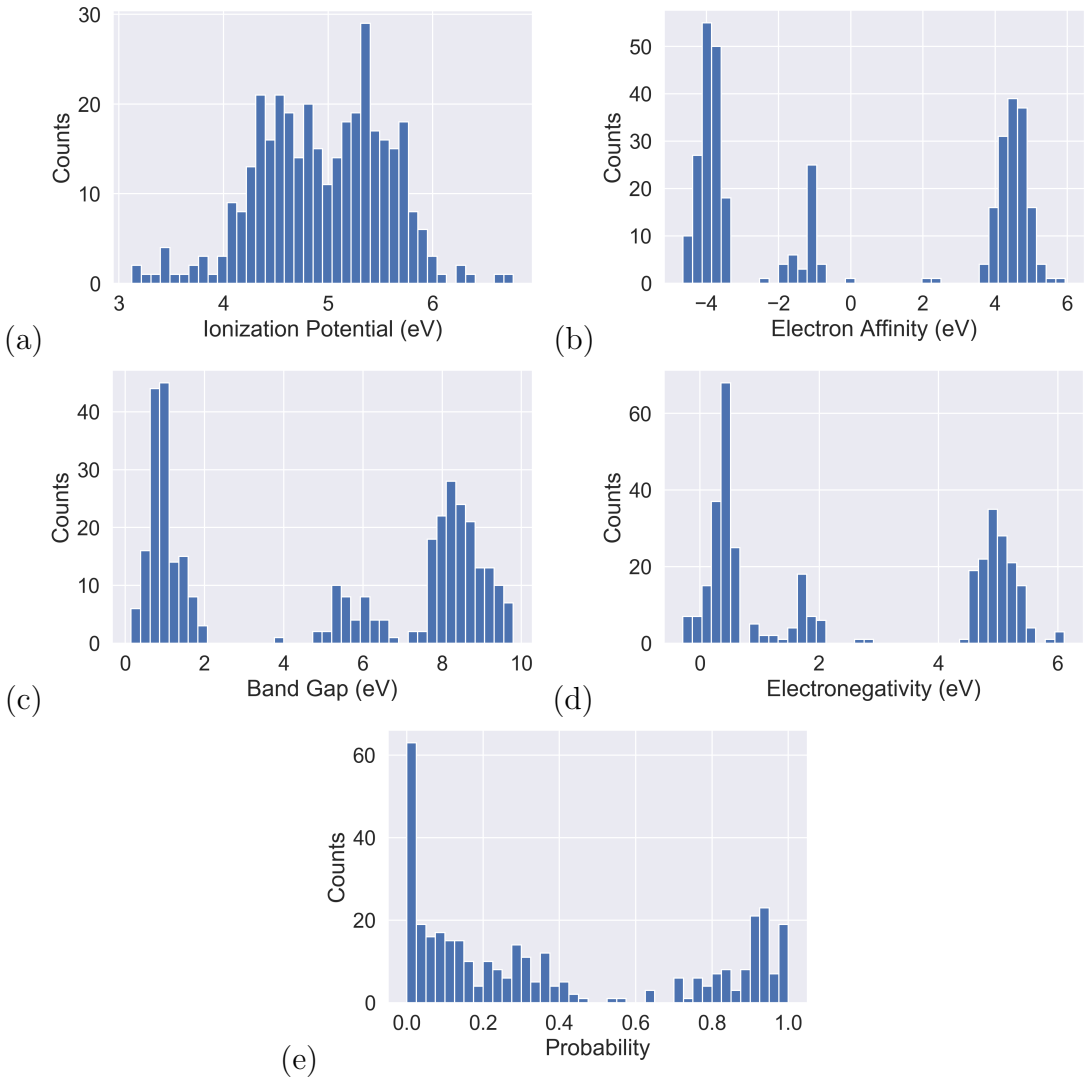

Figure S3: Distribution of property labels within the 355 particle diamond nanoparticle dataset, including the (a) ionization potential, (b) electron affinity, (c) electronic band gap, (d) electronegativity, and (e) the probability of observation, calculated using a Boltzmann distribution and the enthalpy of formation.

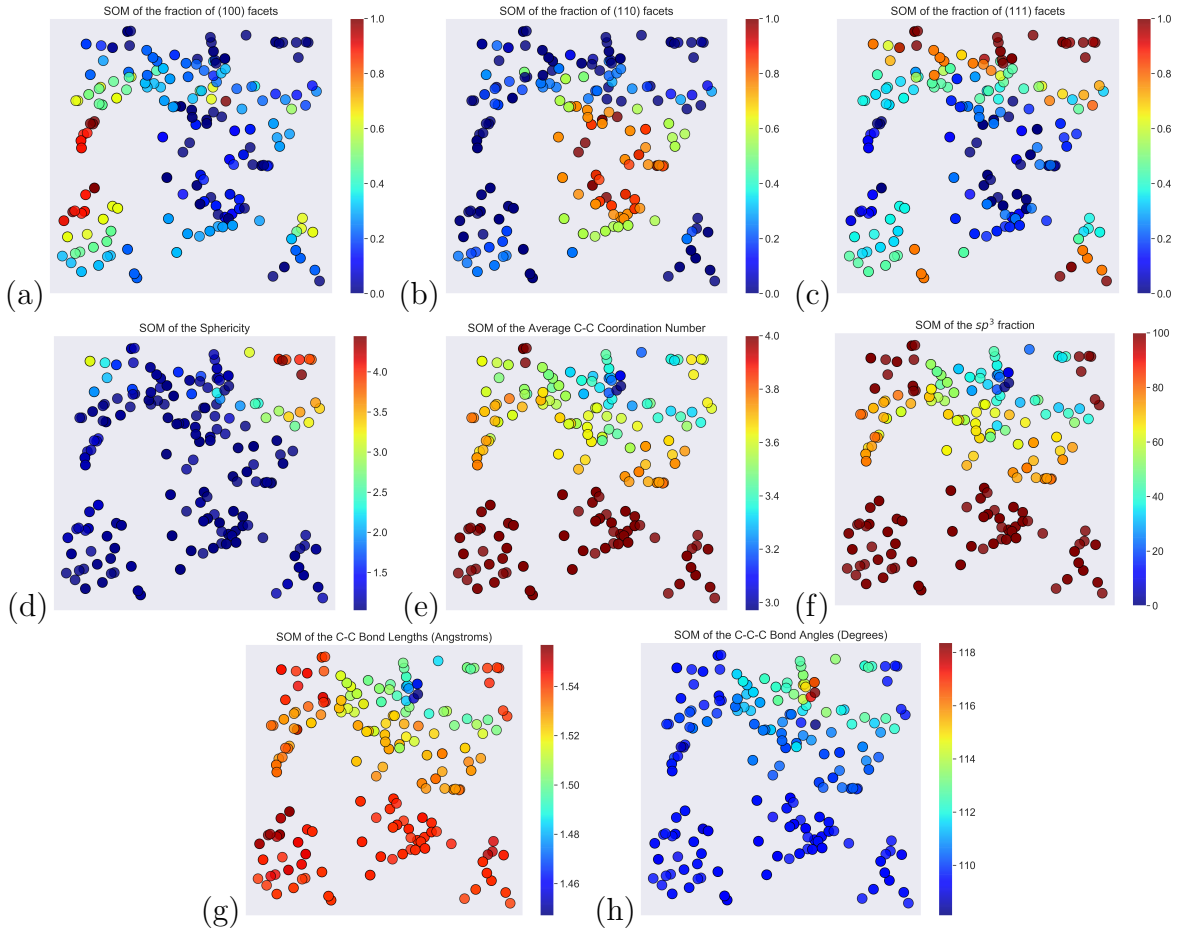

Figure S4: Self-organising maps of structural features within the 355 particle diamond nanoparticle dataset, including the (a) fraction of {100} facet area, (b) fraction of {110} facet area, (c) fraction of {111} facet area, (d) sphericity, (e) average C–C coordination number (4 being perfect diamond, and 3 being perfect graphene), (f) fraction of sp^3 hybridized atoms, (g) average C–C bond length, and (h) the average C–C–C bond angle.

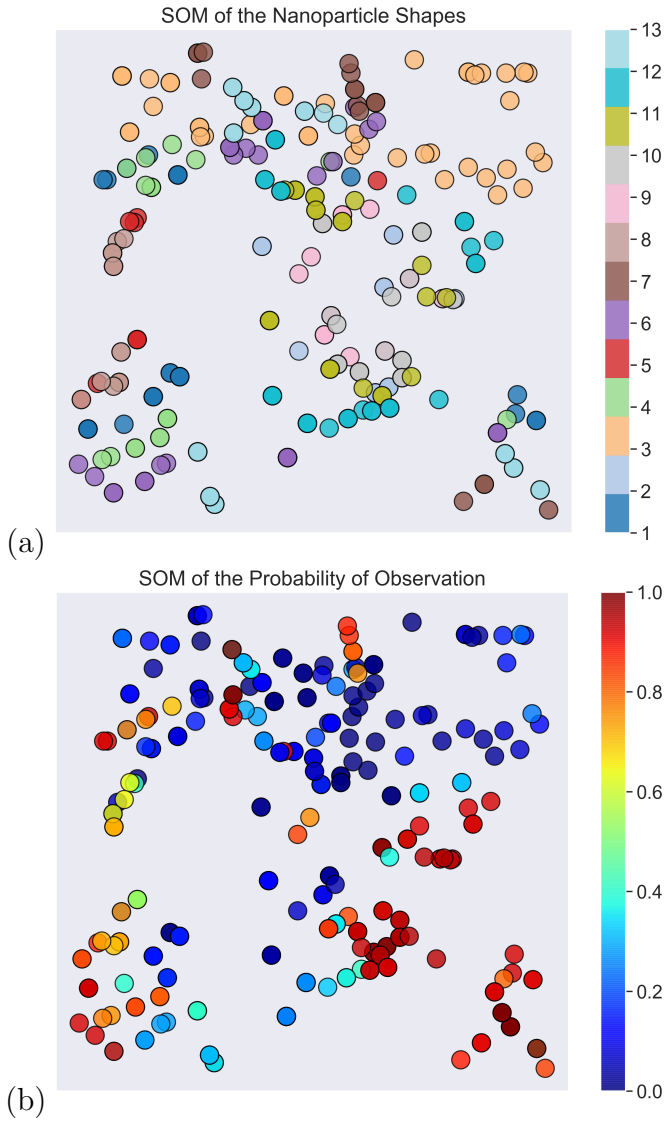

Figure S5: Self-organising maps of property labels within the 355 particle diamond nanoparticle dataset, including (a) the shape, and (b) the probability of observation, calculated using a Boltzmann distribution and the enthalpy of formation. The nanoparticle shapes encoded in (a) are depicted in Figures S6 and Figures S7, below.

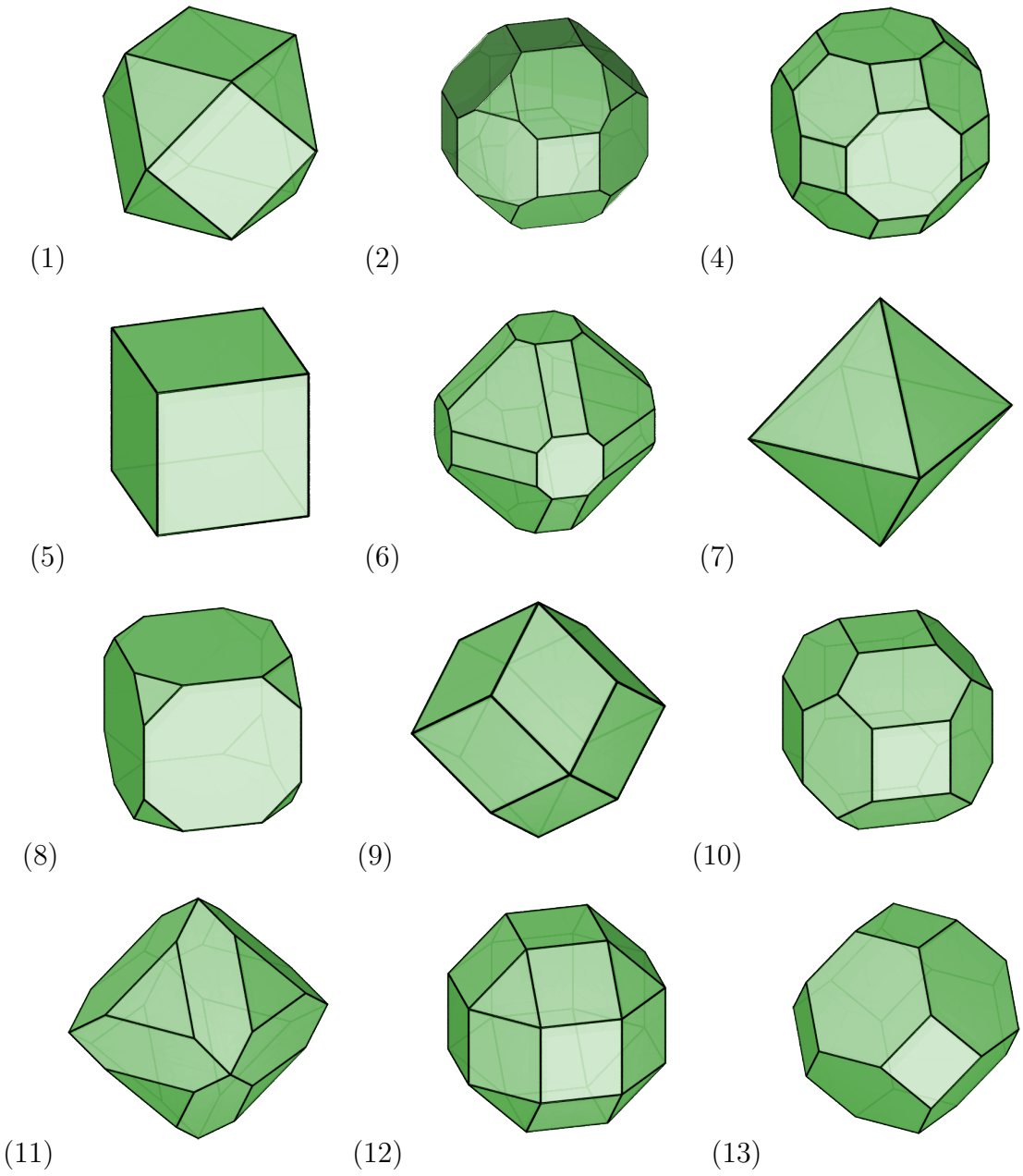

Figure S6: Nanoparticle shapes represented in the dataset used in this study: (1) the cuboctahedron, (2) doubly-truncated rhombic dodecahedron, (4) great rhombicuboctahedron, (5) cube, (6) doubly-truncated octahedron, (7) octahedron, (8) truncated cube, (9) rhombic dodecahedron, (10) rhombi-truncated cube, (11) rhombi-truncated octahedron, (12) small rhombicuboctahedron, (13) truncated octahedron.

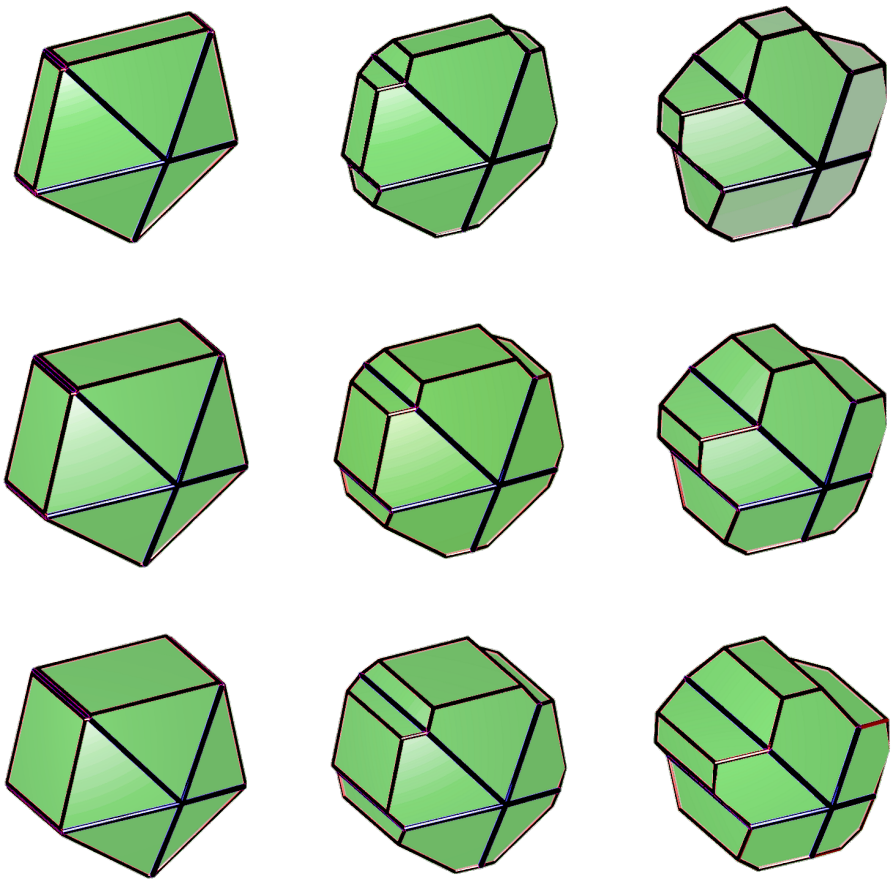

Figure S7: Twinned nanoparticles shapes represented in the dataset used in this study, collectively included in shape category 3.