

## Seturi de date

Sarcina de optimizare a tăieturilor constă în găsirea tăieturilor rectangulare optime date o placă și o listă de defecte. Prin tăieturi *optime* înțelegem că tăieturi care (i) acoperă cât mai mult din placă și (ii) îndeplinesc o serie de constrângeri pre-specificate de utilizator (de exemplu, dimensiune minimă sau număr maxim de defecte conținute de fiecare tăietură).

Data această formulare abstractă a problemei, observăm că putem lucra pe date de intrare relativ simple: o placă poate fi reprezentată ca un dreptunghi, iar defectele ca o listă de dreptunghiuri (de dimensiuni posibil mai mici decât placa). Din acest motiv, am considerat două tipuri de date: sintetice și naturale. Datele sintetice se referă la faptul că acestea generate în mod artificial și aleator, iar datele naturale sunt extrase din baze de date cu imagini naturale și adnoate. Aceste două tipuri sunt descrise în continuare, fiecare în parte.

### Date sintetice

Datele sintetice constă în generarea aleatoare de intrări (placă și defecte) pentru problema de optimizare. Cum placa și defectele sunt reprezentate ca dreptunghiuri, procesul rezultă în generarea a  $n + 1$  dreptunghiuri; acestea sunt obținute astfel: - alegem dimensiunea plăcii (lățime  $h$  și lungime  $w$ ) - alegem un număr  $n$  de defecte - pentru fiecare defect  $i$ , alegem o poziție pe placă  $(x_i, y_i)$  și dimensiunea sa (lățime  $h_i$  și lungime  $w_i$ ).

Pentru fiecare din aceste caracteristici putem alege diferite distribuții. În primele experimente am considerat următoarele distribuții:

$$\begin{array}{lll} w \sim \mathcal{N}(w_{\text{FAS}}, \sigma^2) & h \sim \mathcal{N}(h_{\text{FAS}}, \sigma^2) & \text{dimensiuni placă} \\ n \sim \text{Poi}(7) & & \text{număr defecte} \\ x_i \sim \mathcal{U}(0, h) & y_i \sim \mathcal{U}(0, w) & \text{locație defect} \\ w_i \sim \mathcal{N}\left(\frac{h}{5}, \frac{h^2}{10^2}\right) & h_i \sim \mathcal{N}\left(\frac{h}{10}, \frac{h^2}{10^2}\right) & \text{dimensiune defect} \end{array}$$

unde -  $\mathcal{N}(\mu, \sigma^2)$  reprezintă o distribuție normală (Gaussiană) cu medie  $\mu$  și varianță  $\sigma^2$  -  $\mathcal{U}(a, b)$  reprezintă o distribuție uniformă pe intervalul  $[a, b]$ . -  $\text{Poi}(\lambda)$  reprezintă o distribuție de tip Poisson cu medie  $\lambda$ .

Distribuția dimensiunii plăcii e inspirată din manulalul evaluatorului NHLA: mai precis, am ales media dimensiunii plăci să corespundă gradului *first and second* (FAS), cel mai bun grad ce poate fi acordat unei plăci.

În figura 1, sunt reprezentate exemple de plăci generate conform distribuției definite mai sus.

Avantajele datelor sintetice sunt că acestea sunt ușor de obținut și pot fi generate cantități infinite. Principalul dezavantaj al acestui tip de date e că distribuțiile

utilizate pot fi prea diferite față de datele reale.

## **Date naturale**

Prin date naturale înțelegem faptul că informațiile despre placă și defecte din seturi sunt extrase din date naturale, bazate pe imagini și defecte adnotate manual. În particular, noi am folosit baza de date oferită de cei de la Universitatea din Oulu. Aceasta conține 418 plăci fiecare având patru poze (corespunzând celor patru părți: sus, jos, stânga, dreapta) și adnotată cu 41 tipuri de defecte sub formă de dreptunghiuri.

Câteva imagini din setul lor de date sunt prezentate în figura 2, iar în urma prelucrării renunțăm la informația de aspect și rămânem doar cu defectele și poziționarea lor (vezi figura 3).

Avantajele și dezavantajele acestui tip de date sunt în oglindă față de ce am observat la datele de tip sintetic. Avantajul este că distribuția datelor (dimensiunea plăcilor, locația și dimensiuna defectelor) e probabil mai apropiată de ce vom întâlni în practică. Dezavantajul este cantitatea mai redusă de plăci.