

Detekcija anomalija korištenjem difuzijskih preslikavanja

Doris Đivanović i Karlo Gjogolović

PMF - MO ZG

14. svibnja 2025.

- 1 Teorijski uvod: Difuzijska preslikavanja
 - Uvod - difuzijska preslikavanja
 - Proširenje funkcija izvan uzorka
- 2 Primjena: Detekcija anomalija i konkretni podaci
- 3 Višerezolucijski (Multiscale) pristup
- 4 Zaključak

Uvod - difuzijska preslikavanja

- Analiza podataka - podaci velikih dimenzija \implies nepraktično
- \implies reprezentirati ih podacima iz prostora puno manje dimenzije
- \implies **preslikavanje** u prostor što je moguće manje dimenzije

- Problemi klasteriranja - poseban slučaj: problem **detekcije anomalija** u podacima
- \implies ključni koncepti: povezanost, sličnost, bliskost podataka
- \implies želimo preslikavanje koje čuva te odnose
- \implies tj. takvo da je:

difuzijska udaljenost orig. podataka = euklidska udaljenost njihovih slika

$$D(x, y) = \|\Psi(x) - \Psi(y)\|_2$$

- \implies tzv. **difuzijsko preslikavanje**

Uvod - difuzijska preslikavanja

- $\Gamma = \{x_1, \dots, x_n\}$ skup podataka iz prostora velike dimenzije
- Kako definirati **difuzijsku udaljenost**?
- Kako definirati **difuzijsko preslikavanje**?

Konstrukcija težinskog grafa

- $\Gamma = \{x_1, \dots, x_n\}$ skup podataka iz prostora velike dimenzije
- \implies težinski graf:
 - čvorovi = podaci x_i
 - težina brida $w(x_i, x_j) = \textbf{mjera sličnosti (afiniteta)}$ između x_i i x_j
- \implies za težinsku funkciju w želimo da je:

- simetrična:

$$w(x_i, x_j) = w(x_j, x_i)$$

- nenegativna:

$$w(x_i, x_j) \geq 0$$

- (često) normalizirana:

$$w(x_i, x_j) \in [0, 1], \quad w(x_i, x_i) = 1.$$

- **Kako definirati w ?**

- Prirodno je mjeru afiniteta povezati s euklidskom udaljenošću između podataka i to:

manja udaljenost \iff veća sličnost

- \implies

$$w(x_i, x_j) = \frac{1}{f(\|x_i - x_j\|_2)}, \quad f > 0$$

je simetrična i nenegativna

- \implies uzeti da eksponencijalno ovisi o euklidskoj udaljenosti

- \implies

$$w(x_i, x_j) = \frac{1}{e^{\|x_i - x_j\|_2^2}} = e^{-\|x_i - x_j\|_2^2}$$

- \implies očito je i:

$$w(x_i, x_j) \in [0, 1], \quad w(x_i, x_i) = 1.$$

Funkcija težine: Gaussova jezgra

- Bilo bi još dobro težinsku funkciju skalirati nekim faktorom $\sigma > 0$
- varijabilni parametar \implies promatrati njegov utjecaj s obzirom na njegove različite vrijednosti
- \implies

$$w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|_2^2}{\sigma}},$$

- odnosno, ako želimo označiti da oba podatka doprinose sa $\sigma > 0$:

$$w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|_2^2}{\sigma^2}}$$

- \implies tzv. **Gaussova (gaussovsk) jezgra**
- **Kako odabrati σ ?**

Funkcija težine: Gaussova jezgra

$$w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|_2^2}{\sigma^2}} \quad \sigma > 0$$

- Vrijednost w proporcionalna vrijednosti σ
- σ jako mali \implies većina težina u grafu jako mala, tj. graf jako nepovezan \implies podaci bliski samo sebi
- σ jako velik \implies većina težina u grafu jako velika \implies gotovo svi podaci međusobno povezani, jako bliski
- \implies **OVO JE LOŠE ZA PRIMJENE!**
- Parametar $\sigma > 0$ očito kontrolira (povećava ili umanjuje) utjecaj euklidske udaljenosti dvaju podataka na mjeru njihove sličnosti
- \implies odrediti ga empirijski tako da umanjuje ili uvećava onda kada je prikladno

Funkcija težine: Gaussova jezgra

$$w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|_2^2}{\sigma^2}} \quad \sigma > 0$$

- 1. rješenje: "usrednjenje"

Npr. $\sigma =$

- medijan
- standardna devijacija
- ...

svih međusobnih udaljenosti u grafu.

- **Prednost:** "globalan" σ ...
- **Mana:** ...

Funkcija težine: Gaussova jezgra

- 2. rješenje: skalirati težinu između dva podatka nekim faktorom koji ovisi o bas tim konkretnim podacima
- Želimo "difuzijsku udaljenost" \implies želimo da mjera sličnosti oslikava lokalnu geometriju
- \implies definirati preslikavanje koje svakom podatku pridružuje faktor koji ovisi samo o njegovoj maloj okolini, a ne cijelom skupu podataka:

$$x_i \longrightarrow \sigma_i$$

- \implies svaki od dva podatka doprinosi međusobnoj težini svojim faktorom:

$$w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|_2^2}{\sigma_i \sigma_j}} \quad \sigma_i, \sigma_j > 0$$

Gaussova jezgra za detekciju anomalije

- Primjer - problem **detekcije anomalije**
- Pretpostavljamo:
- Podaci koji nisu anomalija, odnosno u tom smislu "normalni" podaci - nalaze se u "gusto naseljenom" području, tj. nalaze se blizu većine ostalih podataka iz skupa podataka
- Podaci koji jesu anomalija - nalaze se u "rijetko naseljenom" području, odnosno daleko od gotovo svih ostalih podataka iz skupa podataka

Gaussova jezgra za detekciju anomalije

- Primjer - posebno prikladan za problem **detekcije anomalije**:

$$w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|_2^2}{\sigma_i \sigma_j}}$$

$$\sigma_i = \|x_i - x_K\|_2^2$$

- x_K - K -ti najbliži susjed podatka x_i
- Moguća interpretacija ...

Slučajna šetnja i matrica prijelaza

Na temelju težina $w(x_i, x_j)$ definira se slučajna šetnja na skupu podataka.

- Normalizacija - definiramo novu funkciju:

$$p(x_i, x_j) = \frac{w(x_i, x_j)}{d(x_i)}, \quad \text{gdje je } d(x_i) = \sum_{x_j \in \Gamma} w(x_i, x_j).$$

- \implies očito vrijedi:

$$\sum_{x_j \in \Gamma} p(x_i, x_j) = 1, \quad p(x_i, x_j) \geq 0$$

- \implies vrijednost $p(x_i, x_j)$ možemo shvatiti kao vjerojatnost prijelaza iz podatka x_i u podatak x_j u jednom koraku
- \implies matrica P s elementima $p_{ij} = p(x_i, x_j)$ je retčano stohastička matrica prijelaza Markovljeva lanca
- \implies matrica P^t je matrica prijelaza u t koraka, ima elemente $p_{ij}^t = p_t(x_i, x_j)$ - predstavljaju vjerojatnosti prijelaza iz x_i u x_j u t koraka

Slučajna šetnja i matrica prijelaza

- Znamo da retčano stohastička matrica P ima potpun skup svojstvenih vrijednosti, od kojih je jedna i najveća $\lambda = 1$, čiji je svojstveni potprostor skup svih višekratnika jediničnog vektora.
- Spektar matrice P brzo opada, dakle vrijedi:

$$1 = |\lambda_0| \geq |\lambda_1| \geq \dots$$

- Ako s D označimo dijagonalnu matricu čiji su dijagonalni elementi redom

$$d(x_i) = \sum_{x_j \in \Gamma} w(x_i, x_j),$$

matricu P očito možemo prikazati kao:

$$P = D^{-1}W, \quad w_{ij} = w(x_i, x_j).$$

- Matrica težina W standardno se naziva **matrica sličnosti (afiniteta)**.

Spektar matrice prijelaza i difuzijsko preslikavanje

- Matrica P je slična simetričnoj matrici

$$P_s = D^{-1/2} W D^{-1/2},$$

gdje je

$$D^{-1/2} D^{-1/2} = D^{-1},$$

pa te matrice imaju iste svojstvene vrijednosti.

- Kako je matrica P_s simetrična, ima ortonormirani skup lijevih svojstvenih vektora
- Možemo ga dobiti spektralnom dekompozicijom matrice P_s :

$$P_s = U \Lambda U^T.$$

- Za ortonormirani skup lijevih i desnih svojstvenih vektora, Markovljev lanac može se predstaviti ovako:

$$p_t(x_i, x_j) = \sum_{l \geq 0} \lambda_l^t u_l(x_i) v_l(x_j)$$

Spektar matrice prijelaza i difuzijsko preslikavanje

- \implies za ortonormirani skup lijevih svojstvenih vektora, očito bismo svakom podatku x_i mogli pridružiti sljedeće:

$$x_i \longrightarrow (\lambda_0 u_0(x_i), \lambda_1 u_1(x_i), \lambda_2 u_2(x_i), \dots, \lambda_n u_n(x_i))^T.$$

- Kako je u_0 konstantan, možemo ga izostaviti
- Zbog brzog opadanja spektra svojstvenih vrijednosti (λ_l brzo teže k nuli), za određenu točnost gornje sume dovoljno je uzeti određeni broj vodećih svojstvenih parova, neki $l < n$
- \implies svakom podatku x_i mogli bismo pridružiti sljedeće:

$$x_i \in \Gamma \longrightarrow (\lambda_1 u_1(x_i), \lambda_2 u_2(x_i), \dots, \lambda_\ell u_\ell(x_i))^T \in \mathbb{R}^\ell$$

odnosno:

$$x_i \in \Gamma \longrightarrow (\lambda_1^t u_1(x_i), \lambda_2^t u_2(x_i), \dots, \lambda_\ell^t u_\ell(x_i))^T \in \mathbb{R}^\ell$$

za odabrani broj koraka t .

- \implies ovo je dobar **kandidat za difuzijsko preslikavanje!**

Difuzijska udaljenost u prostoru Γ može se definirati ovako:

$$D_t^2(x_i, x_j) = \sum_{x_j \in \Gamma} \frac{(p_t(x_i, x_j) - p_t(x_i, x_j))^2}{u_0(x_j)}$$

Zašto je ovo dobro uzeti kao difuzijsku udaljenost, dakle za **mjeru bliskosti** u prostoru Γ ?

- Mjeri sličnost distribucija vjerojatnosti podataka x_i i x_j nakon t koraka
- Intuitivno: udaljenost između x_i i x_j je mala, tj. bliskost je velika, ako postoji puno kratkih puteva koji povezuju x_i i x_j
- Robusna na šum jer ovisi o svim putevima duljine t .
- Ovisi o cijelom skupu podataka, za razliku od euklidske

Difuzijsko preslikavanje i difuzijska udaljenost

- Izračun pomoću svojstvenih vektora:

$$D_t^2(x_i, x_j) = \sum_{x_j \in \Gamma} \frac{(p_t(x_i, x_j) - p_t(x_j, x_i))^2}{u_0(x_j)} = \sum_{j \geq 1} \lambda_j^{2t} (u_j(x_i) - u_j(x_j))^2$$

- \implies aproksimacija koristeći prvih ℓ svojstvenih vektora je računski efikasna
- Ako **difuzijsko preslikavanje** definiramo ovako:

$$\Psi_t : x_i \in \Gamma \longrightarrow (\lambda_1^t u_1(x_i), \lambda_2^t u_2(x_i), \dots, \lambda_\ell^t u_\ell(x_i))^T \in \mathbb{R}^\ell$$

ovakva **difuzijska udaljenost** uistinu je jednaka Euklidskoj udaljenosti u prostoru difuzijskog preslikavanja:

$$D_t^2(x_i, x_j) = \|\Psi_t(x_i) - \Psi_t(x_j)\|^2$$

Proširenje funkcija izvan uzorka

- Za velike skupove podataka (npr. slike), računanje difuzijskog preslikavanja za sve podatke je nepraktično
- \implies izračunaj preslikavanje za podskup uzoraka $\Gamma' \subseteq \Gamma$, a zatim proširi ulaganje na sve podatke u Γ .
- Poznate metode: Nyströмова ekstenzija, geometrijske harmonike.
- **Višerezolucijska Laplaceova piramida:**
- Aproksimacija funkcije na svakoj razini piramide l pomoću Gaussove jezgre

$$W_l(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma_l)$$

s progresivno manjim σ_l .

- Normalizirana jezgra:

$$K_l = q_l^{-1} W_l, \quad q_l(x_i) = \sum_j W_l(x_i, x_j).$$

Višerezolucijska Laplaceova piramida

- Laplaceova piramidalna reprezentacija funkcije f definirana je iterativno komponentama s_l :

$$s_0(x_k) = \sum_{i=1}^{n_{\text{subset}}} K_0(x_i, x_k) f(x_i)$$

$$s_l(x_k) = \sum_{i=1}^{n_{\text{subset}}} K_l(x_i, x_k) d_l(x_i), \quad l \geq 1$$

$$\text{gdje je } d_l(x_k) = f(x_k) - \sum_{m=0}^{l-1} s_m(x_k), \quad l \geq 1$$

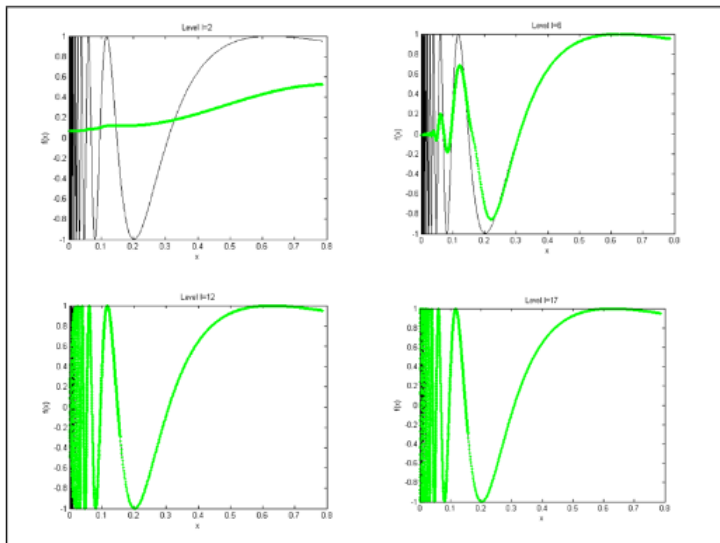
- Iterira se dok norma reziduala ne padne ispod zadanog praga / tolerancije
- Proširenje funkcije f na novu točku $\bar{x}_k \in \Gamma \setminus \Gamma'$:

$$f(\bar{x}_k) = \sum_{l \geq 0} s_l(\bar{x}_k)$$

gdje se $s_l(\bar{x}_k)$ računaju analogno gornjim formulama, koristeći \bar{x}_k umjesto x_k , tj. $K_l(x_i, \bar{x}_k)$

Višerezolucijska Laplaceova piramida

- Primjer: $f(x) = \sin\left(\frac{1}{x+0.1}\right)$



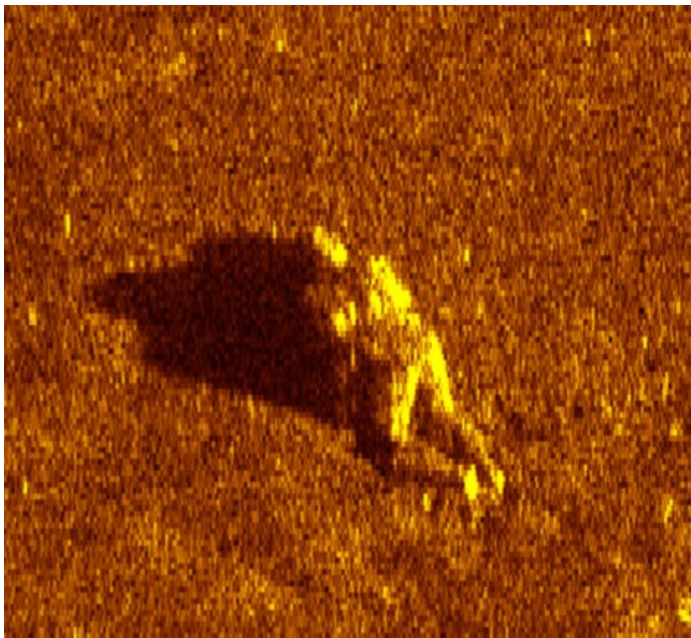
Višerezolucijska Laplaceova piramida

- Provodi se za svaku difuzijsku koordinatu Ψ_j , tj. komponentnu funkciju preslikavanja Ψ , zasebno.

Primjena: Detekcija anomalija na digitalnoj slici

- Cilj: Detektirati dijelove digitalne slike koji odskaku od ostatka (tzv. pozadine), tj. anomalije.
- Problem se svodi na klasteriranje visokodimenzionalnih podataka (u dva klastera)
- Ulazni podaci: digitalna slika (npr. 'grayscale') - podatak tipa *Image*
- \implies Matlab: matrica piksela kao cjelobrojnih vrijednosti intenziteta sive boje $\in [0, 255]$, tj. decimalnih vrijednosti $\in [0, 1]$ nakon normalizacije

Primjena: Detekcija anomalija na digitalnoj slici



Podaci velike dimenzije - *patchevi* digitalne slike

- Podaci za analizu neće nam biti pojedinačni pikseli, već vektorizirani kvadratni isječki (*patchevi*) slike, odabrane dimenzije $patchDim \times patchDim$.
- \implies ovi podaci su vektori u prostoru \mathbb{R}^m , gdje je $m = patchDim^2$.
- dimenziju m smatramo velikom

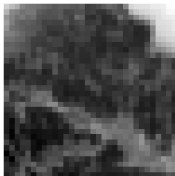
- Dobivaju se izdvajanjem svih preklapajućih blokova (odabrane veličine) iz ulazne slike, tj. matrice
- \implies to se postiže pridruživanjem različitog *patcha* svakom elementu matrice
- \implies uzimati patcheve npr. t.d. da im je element centar ili gornji lijevi vrh
- Može se napraviti *padding* matrice kako bi se mogli izdvojiti *patchevi* centrirani i na rubnim elementima (npr. Matlab funkcija **padarray(...)** s opcijom 'symmetric').
- Za svaki patch, tj. odgovarajuće podatke, pamti se pozicija njegovog centra (ili gornjeg lijevog kuta) u originalnoj slici

Podaci velike dimenzije - *patchevi* digitalne slike

Original Image
with Patch Locations



Patch 1



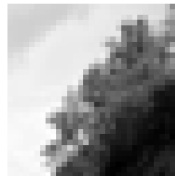
Patch 2



Patch 3



Patch 4



Računanje dif. preslikavanja - Matlab implementacija

- **Uzorkovanje:** zbog računske složenosti, preslikavanje se ne računa na svim patchevima, već na nasumično odabranom podskupu (udio definiran parametrom)
 - MATLAB funkcije: **randperm(...)**, **round()**
 - potencijalni problemi premalenog ili nerelevantog uzorka
- **Matrica sličnosti W za podskup:**
 - Koristi se **lokalno ovisan** σ :

$$\sigma_i = \|x_i - x_{K(i)}\|_2^2$$

- velik skup podataka \implies konstruira se kao **rijetka matrica**: za svaki podatak x_i pronalazi se K **najbližih susjeda**, težine se računaju samo za te parove, zadovoljava se uvjet simetričnosti težina, ostale težine su 0
- MATLAB funkcije: **sparse(...)** za alokaciju, **knnsearch(...)** ili **pdist2(...)** za najbliže susjede

Računanje dif. preslikavanja - Matlab implementacija

- **Simetrična matrica** $P_s = D^{-1/2}WD^{-1/2}$ - dijagonalna matrica $D^{-1/2}$ računa se s pomoću **spdiags(...)**
- **Spektralna dekompozicija** P_s : funkcijom **eigs(...)** dobiva se željeni broj vodećih svojstvenih parova (željena niska dimenzija postavljena kao parametar *maxDim*)
- **Koordinate difuzijskog preslikavanja** - egzaktna za podskup: skaliranjem dobivenih svojstvenih vektora odgovarajućim svojstvenim vrijednostima.
- **Proširenje** na sve originalne podatke (*patches*): izravno pomoću gore opisane Laplaceove piramide, ali također s **rijetkim matricama težina** koristeći k **najbližih susjeda** !

Primjena difuzijskog preslikavanja

Kako koristimo dobiveno difuzijsko preslikavanje za detekciju anomalije među podacima?

- Ako je podatak blizu/sličan drugim podacima - nije anomalija
- Ako je podatak daleko/nije sličan od ostalih podataka - anomalija je
- Računa se afinitet $\bar{w}(i, j)$ između x_i i x_j , za sve x_j u kvadratnom "prozoru" slike oko x_i , koristeći difuzijsku udaljenost i Gaussovu jezgru s globalnim faktorom $\bar{\sigma}$:

$$\bar{w}(i, j) = \exp \left(-\frac{\|\Psi(x_i) - \Psi(x_j)\|^2}{\bar{\sigma}} \right)$$

- Globalni faktor $\bar{\sigma}$ se određuje empirijski, npr. $\bar{\sigma} = r \cdot \sigma_{pair}^2$, gdje je σ_{pair}^2 varijanca difuzijskih udaljenosti nasumično odabranih parova podataka (npr. $r = 20$).
- **Maskiranje:** Unutarnji dio prozora N_i se često isključuje da se izbjegne usporedba s neposrednim susjedima (ako je anomalija veća od jednog patcha). Veličina prozora W i maske M^{mask} ovise o očekivanoj veličini anomalije.

Stupanj anomalije (Anomaly Score)

- **Stupanj anomalije (anomaly score) $C(i)$ za podatak x_i :**

$$C(i) = 1 - \frac{1}{m} \sum_{j \in N_i} \bar{w}(i, j)$$

gdje je N_i skup m susjeda u prozoru oko x_i .

- mala sličnost s okolinom u dif. prostoru \implies visok $C(i)$ (blizu 1) \implies potencijalna anomalija

"Jednoredolucijski pristup" u analizi digitalnih slika

- Problemi ...

Višerezolucijski pristup - Generalna ideja

Umjesto primjene algoritma na samo jednoj rezoluciji slike, koristi se **Gaussova piramida** slike: $\{G_I\}_{I=0}^L$, gdje je G_0 originalna slika (najfinija rezolucija), a G_L slika najgrublje rezolucije.

- Algoritam se odvija iterativno po razinama piramide, počevši od najgrublje ($I = L$) prema najfinijoj ($I = 0$).
- **Ključna ideja:** Informacije o "sumnjivim" podacima s grubljih razina koriste se za usmjeravanje uzorkovanja i analize na finijim razinama

- **Prednosti:**

- Povećava vjerojatnost da će anomalija biti prikladno zastupljena u uzorku na finijim razinama.
- Dozvoljava postavljanje nižih pragova za detekciju sumnjivih područja na grubljim razinama, jer konačna odluka pada na najfinijoj razini.
- Računski efikasniji jer se na grubljim razinama koriste manje značajke (manji isječci/patchevi).

Gaussova piramida

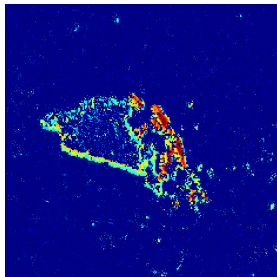
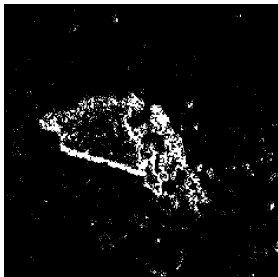
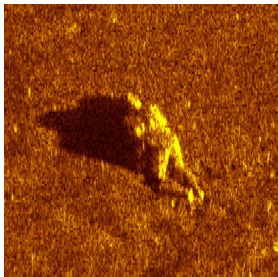
Umjesto primjene algoritma na samo jednoj rezoluciji slike, koristi se **Gaussova piramida** slike: $\{G_l\}_{l=0}^L$, gdje je G_0 originalna slika (najfinija rezolucija), a G_L slika najgrublje rezolucije.

- Gaussov filter
- "downsampling"
- `impyramid(...)`

Gaussova piramida



Rezultat



- Difuzijska preslikavanja su moćan alat za nelinearnu redukciju dimenzionalnosti i analizu strukture podataka.
- Korištenjem patcheva umjesto pojedinačnih piksela omogućuje se hvatanje lokalnog konteksta i strukture, što je bitno za detekciju anomalija u slikama.
- Proširenje funkcija (npr. Laplaceovom piramidom) omogućuje primjenu na velike skupove podataka uzorkovanjem manjeg podskupa za inicijalnu konstrukciju preslikavanja.
- Višerezolucijski (multiscale) pristup dodatno poboljšava robusnost i efikasnost detekcije:
 - Usmjerava proces uzorkovanja na finijim skalama prema područjima koja su se pokazala sumnjivima na grubljim skalama.
 - Povećava šanse za detekciju anomalija različitih veličina.
 - Može biti računski efikasniji od nekih jednerezolucijskih metoda.
- Mjera anomalije temeljena na gustoći susjedstva u prostoru difuzijskih koordinata omogućuje kvantifikaciju "čudnosti" svakog podatkovnog objekta (patcha).