**Algorithm 1** Explicit Shape Regression (ESR)

**Variables**: Training images and labeled shapes $\{I_l, \hat{S}_l\}_{l=1}^{L}$; ESR model $\{R^t\}_{t=1}^{T}$; Testing image $I$; Predicted shape S; *TrainParams*{times of data augment $N_{\text{aug}}$, number of stages $T$};
*TestParams*{number of multiple initializations $N_{\text{int}}$};
*InitSet* which contains exemplar shapes for initialization

***ESRTraining***$(\{I_l, \hat{S}_l\}_{l=1}^{L}, \textit{TrainParams}, \textit{InitSet})$
// augment training data
$\{I_i, \hat{S}_i, S_i^0\}_{i=1}^{N} \leftarrow \textit{Initialization}\ (\{I_l, \hat{S}_l\}_{l=1}^{L}, N_{\text{aug}}, \textit{InitSet})$
**for** $t$ from 1 to $T$
  $Y \leftarrow \{M_{S_i^{t-1}} \circ (\hat{S}_i - S_i^{t-1})\}_{i=1}^{N}$ // compute normalized targets
  $R^t \leftarrow \textit{LearnStageRegressor}(Y, \{I_i, S_i^{t-1}\}_{i=1}^{N})$ // using Eq. (3)
  **for** $i$ from 1 to $N$
    $S_i^t \leftarrow S_i^{t-1} + M_{S_i^{t-1}}^{-1} \circ R^t(I_i, S_i^{t-1})$
**return** $\{R^t\}_{t=1}^{T}$

***ESRTesting***$(I, \{R^t\}_{t=1}^{T}, \textit{TestParams}, \textit{InitSet})$
// multiple initializations
$\{I_i, *, S_i^0\}_{i=1}^{N_{\text{int}}} \leftarrow \textit{Initialization}\ (\{I, *\}, N_{\text{int}}, \textit{InitSet})$
**for** $t$ from 1 to $T$
  **for** $i$ from 1 to $N_{\text{int}}$
    $S_i^t \leftarrow S_i^{t-1} + M_{S_i^{t-1}}^{-1} \circ R^t(I_i, S_i^{t-1})$
$S \leftarrow \textit{CombineMultipleResutls}(\{S_i^T\}_{i=1}^{N_{\text{int}}})$
**return** $S$

***Initialization***$(\{I_c, \hat{S}_c\}_{c=1}^{C}, D, \textit{InitSet})$
$i \leftarrow 1$
**for** $c$ from 1 to $C$
  **for** $d$ from 1 to $D$
    $S_i^o \leftarrow$ sampling an exemplar shape from *InitSet*
    $\{I_i^o, \hat{S}_i^o\} \leftarrow \{I_c, \hat{S}_c\}$
    $i \leftarrow i + 1$
return $\{I_i^o, \hat{S}_i^o, S_i^o\}_{i=1}^{CD}$

**Algorithm 3** Correlation-based feature selection

**Input**: regression targets $Y \in \Re^{N \times 2N_{\text{fp}}}$; shape indexed pixel features $\rho \in \Re^{N \times P}$; pixel-pixel covariance $\text{cov}(\rho) \in \Re^{P \times P}$; number of features of a fern $F$;
**Output**: The selected pixel-difference features $\{\rho_{m_\text{f}} - \rho_{n_\text{f}}\}_{\text{f}=1}^{F}$ and the corresponding indices $\{m_\text{f}, n_\text{f}\}_{\text{f}=1}^{F}$;

***CorrelationBasedFeatureSelection***$(Y, \text{cov}(\rho), F)$
**for** f from 1 to $F$
  $v \leftarrow \text{randn}(2N_{\text{fp}}, 1)$ // draw a random projection from unit Gaussian
  $Y_{\text{prob}} \leftarrow Yv$ // random projection
  $\text{cov}(Y_{\text{prob}}, \rho) \in \Re^{1 \times P} \leftarrow$ compute target-pixel covariance
  $\sigma(Y_{\text{prob}}) \leftarrow$ compute sample variance of $Y_{\text{prob}}$
  $m_\text{f} = 1; n_\text{f} = 1;$
  **for** $m$ from 1 to $P$
    **for** $n$ from 1 to $P$
      $\text{corr}(Y_{\text{prob}}, \rho_m - \rho_n) \leftarrow$ compute correlation using Eq. (11)
      **if** $\text{corr}(Y_{\text{prob}}, \rho_m - \rho_n) > \text{corr}(Y_{\text{prob}}, \rho_{m_\text{f}} - \rho_{n_\text{f}})$
        $m_\text{f} = m; n_\text{f} = n;$
**return** $\{\rho_{m_\text{f}} - \rho_{n_\text{f}}\}_{\text{f}=1}^{F}, \{m_\text{f}, n_\text{f}\}_{\text{f}=1}^{F}$

**Algorithm 2** Shape indexed features

**Variables**: images and corresponding estimated shapes $\{I_i, S_i\}_{i=1}^{N}$; number of shape indexed pixel features $P$; number of facial points $N_{\text{fp}}$; the range of local coordinate $\kappa$; local coordinates $\{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P}$; shape indexed pixel features $\rho \in \Re^{N \times P}$ ; shape indexed pixel-difference features $X \in \Re^{N \times P^2}$;

***GenerateShapeIndexedFeatures***$(\{I_i, S_i\}_{i=1}^{N}, N_{\text{fp}}, P, \kappa)$
$\{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P} \leftarrow \textit{GenerateLocalCoordinates}(\textit{FeatureParams})$
$\rho \leftarrow \textit{ExtractShapeIndexedPixels}(\{I_i, S_i\}_{i=1}^{N}, \{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P})$
$X \leftarrow$ pairwise difference of all columns of $\rho$
**return** $\{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P}, \rho, X$

***GenerateLocalCoordinates***$(N_{\text{fp}}, P, \kappa)$
**for** $\alpha$ from 1 to $P$
  $l_\alpha \leftarrow$ randomly drawn a integer in $[1, N_{\text{fp}}]$
  $\Delta_\alpha^{l_\alpha} \leftarrow$ randomly drawn two floats in $[-\kappa, \kappa]$
**return** $\{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P}$

***ExtractShapeIndexedPixels***$(\{I_i, S_i\}_{i=1}^{N}, \{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P})$
**for** $i$ from 1 to $N$
  **for** $\alpha$ from 1 to $P$
    $\mu_\alpha \leftarrow \pi_{l_\alpha} \circ S_i + M_{S_i}^{-1} \circ \Delta^{l_\alpha}$
    $\rho_{i\alpha} \leftarrow I_i(\mu_\alpha)$
**return** $\rho$

**Algorithm 4** Internal-level boosted regression

**Variables**: regression targets $Y \in \Re^{N \times 2N_{\text{fp}}}$; training images and corresponding estimated shapes $\{I_i, S_i\}_{i=1}^{N}$; training parameters *TrainParams*{$N_{\text{fp}}, P, \kappa, F, K$}; the stage regressor $R$; testing image and corresponding estimated shape $\{I, S\}$;

***LearnStageRegressor***$(Y, \{I_i, S_i\}_{i=1}^{N}, \textit{TrainParams})$
$\{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P} \leftarrow \textit{GenerateLocalCoordinates}(N_{\text{fp}}, P, \kappa)$
$\rho \leftarrow \textit{ExtractShapeIndexedPixels}(\{I_i, S_i\}_{i=1}^{N}, \{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P})$
$\text{cov}(\rho) \leftarrow$ pre-compute pixel-pixel covariance
$Y^0 \leftarrow Y$ // initialization
**for** $k$ from 1 to $K$
  $\{\rho_{m_\text{f}} - \rho_{n_\text{f}}\}_{\text{f}=1}^{F}, \{m_\text{f}, n_\text{f}\}_{\text{f}=1}^{F} \leftarrow$
    $\textit{CorrelationBasedFeatureSelection}(Y^{k-1}, \text{cov}(\rho), F)$
  $\{\theta_\text{f}\}_{\text{f}=1}^{F} \leftarrow$ sample $F$ thresholds from an uniform distribution
  $\{\Omega_b\}_{b=1}^{2^F} \leftarrow$ partition training samples into $2^F$ bins
  $\{y_b\}_{b=1}^{2^F} \leftarrow$ compute the outputs of all bins using Eq. (7)
  $r_k \leftarrow \{\{m_\text{f}, n_\text{f}\}_{\text{f}=1}^{F}, \{\theta_\text{f}\}_{\text{f}=1}^{F}, \{y_b\}_{b=1}^{2^F}\}$ // construct a fern
  $Y^k \leftarrow Y^{k-1} - r^k(\{\rho_{m_\text{f}} - \rho_{n_\text{f}}\}_{\text{f}=1}^{F})$ // update the targets
$R \leftarrow \{\{r^k\}_{k=1}^{K}, \{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P}\}$ // construct stage regressor
**return** $R$

***ApplyStageRegressor***$(I, S, R)$ // i.e. $R(I, S)$
$\rho \leftarrow \textit{ExtractShapeIndexedPixels}(\{I, S\}, \{\Delta_\alpha^{l_\alpha}\}_{\alpha=1}^{P})$
$\delta S \leftarrow \mathbf{0}$
**for** $k$ from 1 to $K$
  $\delta S \leftarrow \delta S + r^k(\{\rho_{m_\text{f}} - \rho_{n_\text{f}}\}_{\text{f}=1}^{F})$
**return** $\delta S$