

## Задание 2. EM алгоритм для детектива.

Медведев Алексей Владимирович

# Теория

Формулы:

$$1 \quad p(d_k|X_k, \theta, A) = \frac{p(X_k|\theta, d_k)p(d_k|A)}{\sum_{d^h} \sum_{d^w} p(X_k|\theta, d)p(d|A)}$$

$$2.1 \quad \bullet \quad J = \mathbb{E}_{q(d)} \log p(X, d|\theta, A) = \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} \left( \underbrace{\log p(X_k|\theta, d_k)}_{const(A)} + \underbrace{\log p(d_k|A)}_{const(F, B, s)} \right) \rightarrow J = J_A + J_{F, B, s}$$

$$\bullet \quad J_A = \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} \underbrace{\log p(d_k|A)}_{p_k} = \sum_{k=1}^{\mathbf{K}} \mathbb{E}_{q(d)} p_k = \left[ q(d) = \prod_{i=1}^{i=\mathbf{K}} q_i(d_i) \right] =$$

$$= \sum_{k=1}^{\mathbf{K}} \sum_{d_1} \dots \sum_{d_{\mathbf{K}}} q_1(d_1) \cdot \dots \cdot q_{\mathbf{K}}(d_{\mathbf{K}}) p_k = \sum_{k=1}^{\mathbf{K}} \sum_{d_k} q_k(d_k) p_k$$

$$\bullet \quad \frac{dJ_A}{dA_{i,j}} = \underbrace{\sum_{k=1}^{\mathbf{K}} q_k((i, j))}_{q_{i,j}} \frac{1}{A_{i,j}}$$

$$\bullet \quad L(A, \mu, \lambda) = J + \lambda \left( 1 - \sum_i \sum_j A_{i,j} \right) + \sum_i \sum_j \mu_{i,j} A_{i,j}$$

•

$$\left\{ \begin{array}{l} \frac{dL}{dA_{i,j}} = \frac{q_{i,j}}{A_{i,j}} - \lambda + \mu_{i,j} = 0 \\ \sum_i \sum_j A_{i,j} = 1 \\ A_{i,j} \geq 0 \\ \mu_{i,j} A_{i,j} = 0 \end{array} \right.$$

$$\bullet \quad \text{Если } A_{i,j} \neq 0, \text{ то } \mu_{i,j} = 0: A_{i,j} = \frac{q_{i,j}}{\lambda} \rightarrow \sum_i \sum_j \frac{q_{i,j}}{\lambda} = 1 \rightarrow \lambda = \mathbf{K} \rightarrow A_{i,j} = \frac{q_{i,j}}{\mathbf{K}}$$

2.2 •

$$J_{F, B, s} = \mathbb{E}_{q(d)} \sum_{i=1}^{\mathbf{K}} \log p(X_k|\theta, d_k) = \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} \underbrace{\left[ \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} \log \mathcal{N}(\mathbf{X}_k(i, j)|\mathbf{B}(i, j), s^2) \right]}_{const(F)} +$$

$$+ \left[ \sum_{i=0}^{\mathbf{h}} \sum_{j=0}^{\mathbf{w}} \underbrace{\log \mathcal{N}(\mathbf{X}_k(i + d_k^h, j + d_k^w)|\mathbf{F}(i, j), s^2)}_{const(B)} - \underbrace{\log \mathcal{N}(\mathbf{X}_k(i + d_k^h, j + d_k^w)|\mathbf{B}(i + d_k^h, j + d_k^w), s^2)}_{const(F)} \right]$$

$$\bullet \quad J_F = \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{h}} \sum_{j=0}^{\mathbf{w}} \log \mathcal{N}(\mathbf{X}_k(i + d_k^h, j + d_k^w)|\mathbf{F}(i, j), s^2)$$

$$\bullet \quad \frac{dJ_F}{dF_{i,j}} = \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} - \frac{\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{F}(i, j)}{s^2} = 0 \rightarrow \mathbf{F}_{i,j} = \sum_{k=1}^{\mathbf{K}} \sum_{d_k} \frac{\mathbf{X}_k(i + d_k^h, j + d_k^w) q_k(d_k)}{\mathbf{K}}$$

2.3 •

$$\begin{aligned}
J_B &= \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} \log \mathcal{N}(\mathbf{X}_k(i, j) | \mathbf{B}(i, j), s^2) - \\
&\quad - \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{h}} \sum_{j=0}^{\mathbf{w}} \log \mathcal{N}(\mathbf{X}_k(i + d_k^h, j + d_k^w) | \mathbf{B}(i + d_k^h, j + d_k^w), s^2) = \\
&= \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} \left[ \mathbb{E}_{q(d)} \log \mathcal{N}(\mathbf{X}_k(i, j) | \mathbf{B}(i, j), s^2) - \sum_{n=0}^{\mathbf{h}} \sum_{m=0}^{\mathbf{w}} \log \mathcal{N}(\mathbf{X}_k(i, j) | \mathbf{B}(i, j), s^2) q_k((i - n, j - m)) \right] = \\
&= \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} \log \mathcal{N}(\mathbf{X}_k(i, j) | \mathbf{B}(i, j), s^2) \left[ 1 - \sum_{n=0}^{\mathbf{h}} \sum_{m=0}^{\mathbf{w}} q_k((i - n, j - m)) \right] \\
\bullet \quad \frac{dJ_B}{dB_{i,j}} &= \sum_{k=1}^{\mathbf{K}} - \frac{\mathbf{X}_k(i,j) - \mathbf{B}(i,j)}{s^2} \left[ 1 - \sum_{n=0}^{\mathbf{h}} \sum_{m=0}^{\mathbf{w}} q_k((i - n, j - m)) \right] = 0 \rightarrow B_{i,j} = \frac{\sum_{k=1}^{\mathbf{K}} \mathbf{X}_k(i,j) \left[ 1 - \sum_{n=0}^{\mathbf{h}} \sum_{m=0}^{\mathbf{w}} q_k((i - n, j - m)) \right]}{\sum_{k=1}^{\mathbf{K}} \left[ 1 - \sum_{n=0}^{\mathbf{h}} \sum_{m=0}^{\mathbf{w}} q_k((i - n, j - m)) \right]}
\end{aligned}$$

2.4

•

$$\begin{aligned}
\frac{dJ_{F,B,s}}{ds} &= \mathbb{E}_{q(d)} \sum_{k=1}^{\mathbf{K}} \left[ \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} -\frac{1}{s} + \frac{(\mathbf{X}_k(i,j) - \mathbf{B}(i,j))^2}{s^3} \right] + \\
&\quad + \left[ \sum_{i=0}^{\mathbf{h}} \sum_{j=0}^{\mathbf{w}} \frac{(\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{F}(i,j))^2}{s^3} - \frac{(\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{B}(i + d_k^h, j + d_k^w))^2}{s^3} \right] = \\
&= \frac{-\mathbf{K} \cdot \mathbf{H} \cdot \mathbf{W} s^2 + \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} (\mathbf{X}_k(i,j) - \mathbf{B}(i,j))^2 + \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{h}} \sum_{j=0}^{\mathbf{w}} \sum_{d_k} ((\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{F}(i,j))^2 - (\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{B}(i + d_k^h, j + d_k^w))^2) q_k(d_k)}{s^3} = 0
\end{aligned}$$

$$\bullet \quad s = \sqrt{\frac{\sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} (\mathbf{X}_k(i,j) - \mathbf{B}(i,j))^2 + \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{h}} \sum_{j=0}^{\mathbf{w}} \sum_{d_k} ((\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{F}(i,j))^2 - (\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{B}(i + d_k^h, j + d_k^w))^2) q_k(d_k)}{\mathbf{K} \cdot \mathbf{H} \cdot \mathbf{W}}}$$

В итоге:

$$\bullet \quad \boxed{A_{i,j} = \frac{q_{i,j}}{\mathbf{K}}}$$

$$\bullet \quad \boxed{\mathbf{F}_{i,j} = \sum_{k=1}^{\mathbf{K}} \sum_{d_k} \frac{\mathbf{X}_k(i + d_k^h, j + d_k^w) q_k(d_k)}{\mathbf{K}}}$$

$$\bullet \quad \boxed{B_{i,j} = \frac{\sum_{k=1}^{\mathbf{K}} \mathbf{X}_k(i, j) \left[ 1 - \sum_{n=0}^{\mathbf{h}} \sum_{m=0}^{\mathbf{w}} q_k((i - n, j - m)) \right]}{\sum_{k=1}^{\mathbf{K}} \left[ 1 - \sum_{n=0}^{\mathbf{h}} \sum_{m=0}^{\mathbf{w}} q_k((i - n, j - m)) \right]}}$$

$$\bullet \quad s = \sqrt{\frac{\sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{H}} \sum_{j=0}^{\mathbf{W}} (\mathbf{X}_k(i,j) - \mathbf{B}(i,j))^2 + \sum_{k=1}^{\mathbf{K}} \sum_{i=0}^{\mathbf{h}} \sum_{j=0}^{\mathbf{w}} \sum_{d_k} ((\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{F}(i,j))^2 - (\mathbf{X}_k(i + d_k^h, j + d_k^w) - \mathbf{B}(i + d_k^h, j + d_k^w))^2) q_k(d_k)}{\mathbf{K} \cdot \mathbf{H} \cdot \mathbf{W}}}$$

$$\bullet \quad \boxed{\mathcal{L}(q, \theta, A) = \mathbb{E}_{q(d)} \log p(X, d | \theta, A) - \mathbb{E}_{q(d)} \log q(d)}$$

## Анализ

### Исследование 1

Посмотрим как работает полученный выше ЕМ алгоритм для разных начальных приближений. Далее будут проиллюстрированы начальные данные с шумом и без шума, а также результат работы алгоритма при различных инициализациях матриц **A**, **B**, **F**. О качестве результата работы алгоритма можно судить по полученным изображениям лица, а также по графикам функции потерь.

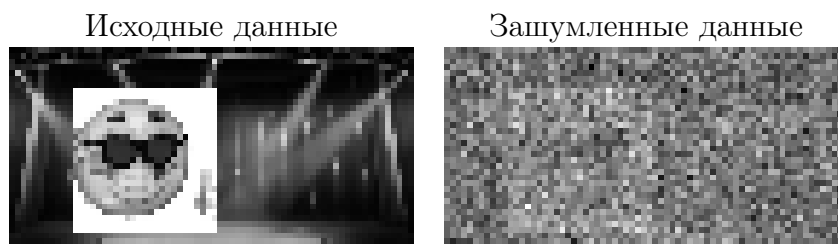


Таблица 1: Визуализация входных данных.

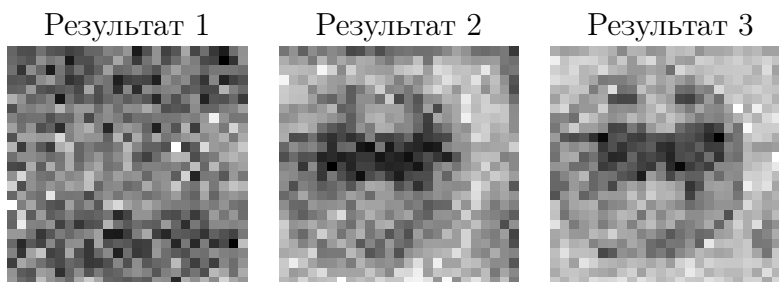


Таблица 2: Визуализация результатов для различных начальных приближений.

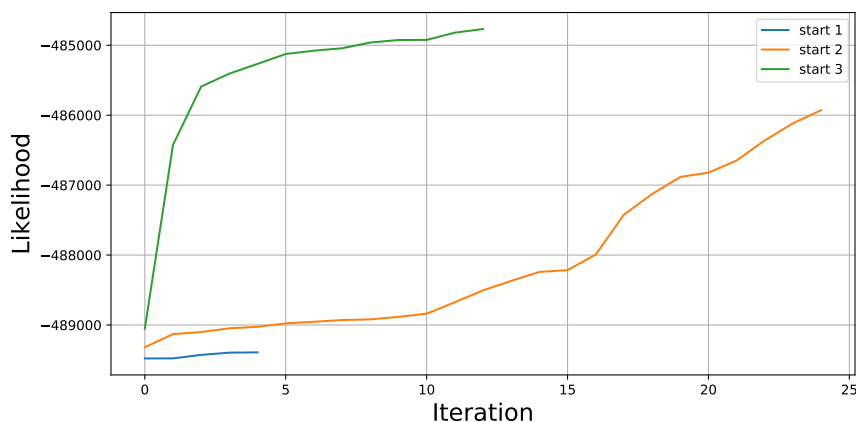


Рис. 1: Нижняя оценка логарифма правдоподобия для различных начальных приближений.

**Вывод:** Для получения лучшего результата, в данной задаче имеет смысл запускать алгоритм из различных начальных приближений. Так как оптимизация идет то по одной части параметров, то по другой, можно сойтись к локальному оптимуму и «увязнуть» там.

## Исследование 2

В этом исследовании мы изучим влияние размеров выборки изображений и её зашумленности на конечный результат алгоритма.

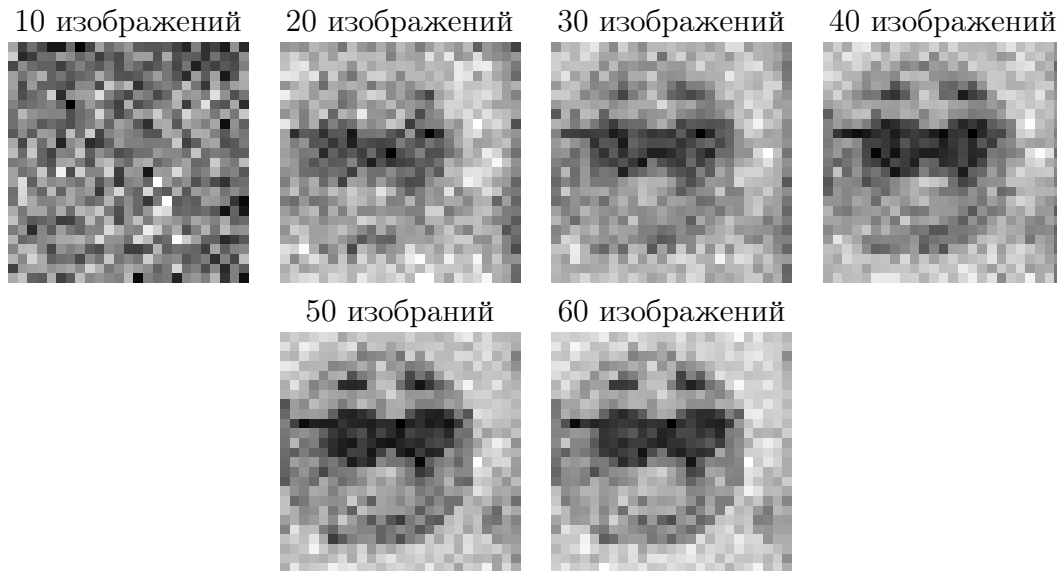


Таблица 3: Визуализация результатов для различных начальных приближений.

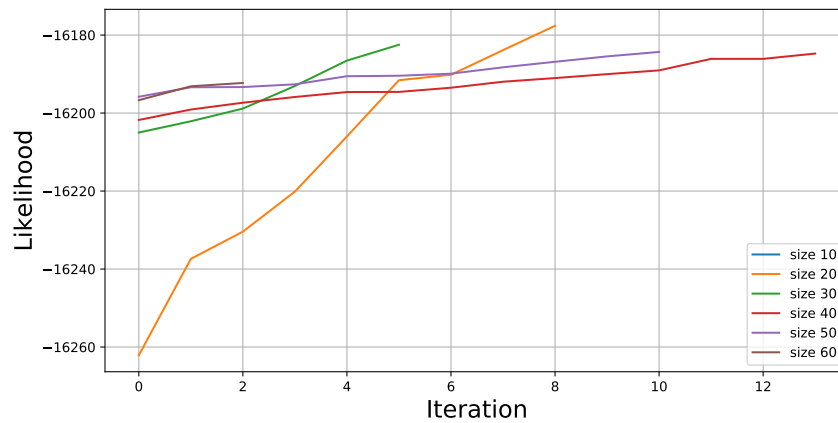


Рис. 2: Нижняя оценка логарифма правдоподобия для выборок разных размеров.

**Вывод:** В этом исследовании мы варьировали только размер выборки, оставив начальное приближение неизменным на всех этапах эксперимента. Из графика изменения нижней оценки на логарифм правдоподобия видно, что чем больше размер выборки, тем эффективнее работает алгоритм уже на первой итерации и тем быстрее сходимость к оптимуму. Визуально качество результата также растет с увеличением выборки.

Рассмотрим теперь влияние уровня шума. Уровень шума в первой части исследования был  $\sigma = 180$ . Теперь будем варьировать уровень шума при фиксированном размере выборки в 200 изображений.

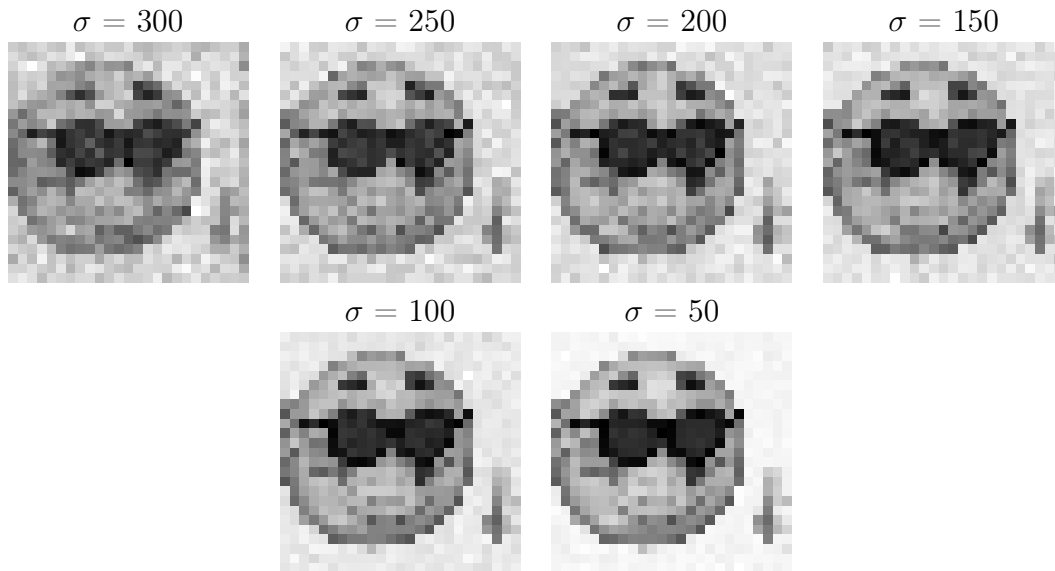


Таблица 4: Визуализация результатов для различных уровней шума.

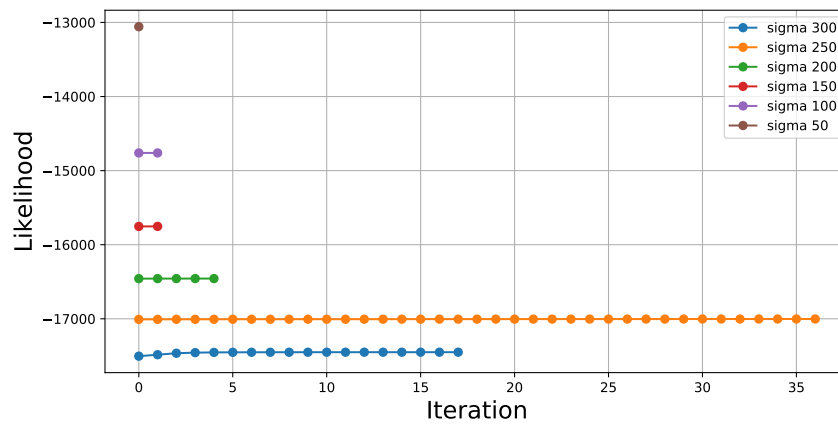


Рис. 3: Нижняя оценка логарифма правдоподобия для разных уровней шума.

**Вывод:** Исходя из результатов первой и второй части исследования ЕМ алгоритм перестает выдавать осмысленные результаты при уровне шума, значительно превышающем размер выборки.

### Исследование 3

В этом исследовании сравним работу ЕМ и hard-ЕМ алгоритмов. Для этого запустим оба алгоритма на наших данных из одного начального приближения.

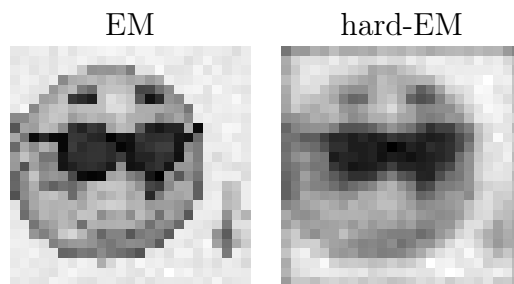


Таблица 5: Результаты работы EM и hard-EM алгоритмов.

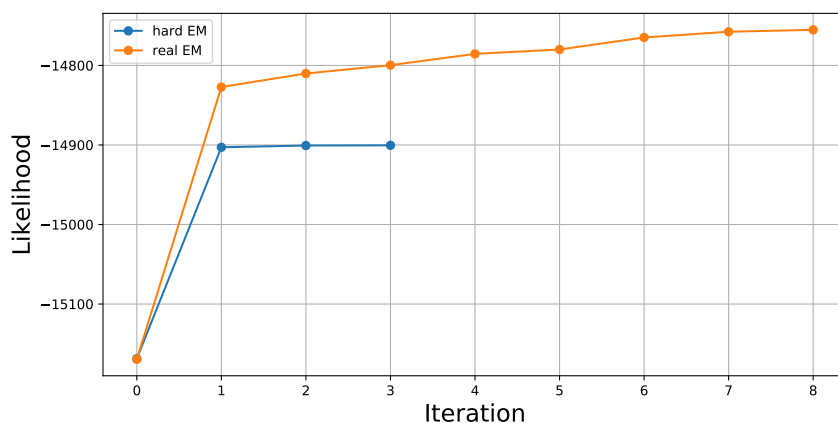


Рис. 4: Нижняя оценка логарифма правдоподобия для EM и hard-EM алгоритмов.

**Вывод:** hard-EM отработал почти в два раза быстрее на наших данных, но результат оказался гораздо хуже, чем у обычного EM. Думаю разница в работе так заметна, потому что второй алгоритм использует для пересчета параметров  $\mathbf{A}, \mathbf{F}, \mathbf{B}, \mathbf{s}$  гораздо больше данных. Каждой такой порции данных обычный EM ставит в соответствие какой-то вес, такая модель гораздо сложнее и имеет больше возможностей для «настройки».

#### Исследование 4

В данном исследовании применим EM алгоритм к зашумленным снимкам преступника.

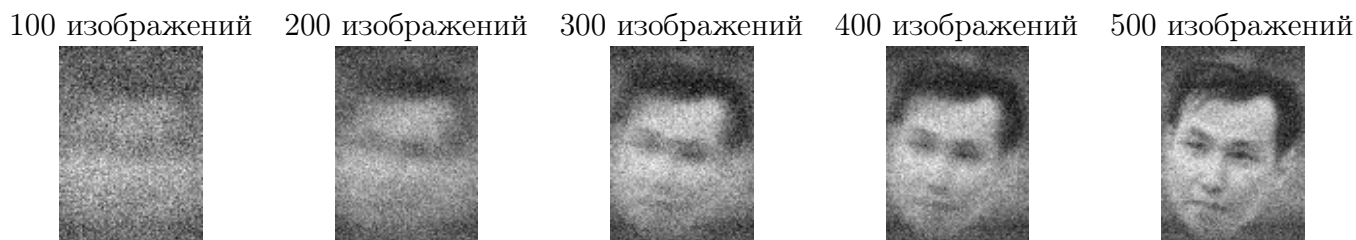


Таблица 6: Визуализация результатов для различных размеров выборки.

**Вывод:** Эффективность алгоритма заметно растет с размером выборки.

## Исследование 5

В качестве улучшения, я бы предложил инициализировать матрицы  $F$  и  $B$  не произвольно, а с помощью результатов, полученных при других начальных приближениях. Это должно ускорить алгоритм.