Практическое задание 2

Королев Кирилл

17 ноября 2023 г.

$$p(X_k \mid d_k, \theta) = \prod_{i,j} \frac{1}{\sqrt{2\pi s^2}} (e^{-\frac{1}{2s^2}(X_k(i,j) - F(i-d_k^h, j-d_k^w))^2})^{[(i,j) \in FA(d_k)]} (e^{-\frac{1}{2s^2}(X_k(i,j) - B(i,j))^2})^{[(i,j) \notin FA(d_k)]}$$

$$p(d_k \mid A) = A(d_k^h, d_k^w)$$

E-шаг

$$q(d_k) = p(d_k \mid X_k, \theta, A) = \frac{p(X_k \mid d_k, \theta) p(d_k \mid A)}{\sum_{\hat{d}_k} p(X_k \mid \hat{d}_k, \theta) p(\hat{d}_k \mid A)}$$

$$q(d) = p(d \mid X, \theta, A) = \prod_k p(d_k \mid X_k, \theta, A)$$

$$q(d) = Q_{mnk} \in \mathbb{R}^{(H-h+1)\times(W-w+1)\times K}$$

M-шаг

$$\mathbb{E}_{q(d)}(\log p(X, d \mid \theta, A)) \to \max_{\theta, A}$$

$$\log p(X, d \mid \theta, A) = \sum_k (\log p(X_k \mid d_k, \theta) + \log p(d_k \mid A))$$

$$\mathbb{E}_{q(d)}(\log p(X, d \mid \theta, A)) = \sum_{m,n,k} Q_{mnk} (\log p(X_k \mid d_k = (m,n), \theta) + \log p(d_k = (m,n) \mid A))$$

Поиск оптимального A

Найдем оптимальное A. Для этого надо решить задачу условной оптимизации.

$$\sum_{m,n} A_{mn} = 1, \quad A_{mn} \geq 0$$

$$L = \sum_{m,n,k} Q_{mnk} (\log p(X_k \mid d_k = (m,n), \theta) + \log p(d_k = (m,n) \mid A)) - \mu(\sum_{m,n} A_{mn} - 1) - \sum_{m,n} \lambda_{mn} A_{mn} \to \max_{A,\mu,\lambda}$$

Запишем условия ККТ.

$$\begin{cases} 0 = \frac{\partial L}{\partial A_{mn}} = \sum_k Q_{mnk} \frac{1}{A_{mn}} - \mu - \lambda_{mn} \\ 0 = \frac{\partial L}{\partial \mu} = \sum_{m,n} A_{mn} - 1 \\ \lambda_{mn} \geq 0 \\ \lambda_{mn} A_{mn} = 0 \\ A_{mn} \geq 0 \end{cases}. \tag{1}$$

$$A_{mn} = \frac{\sum_k Q_{mnk}}{\mu + \lambda_{mn}}$$

Из условия дополняющей нежесткости, если $A_{mn} > 0$, тогда $\lambda_{mn} = 0$. Если $A_{mn} = 0$, тогда $\lambda_{mn} \geq 0$. Для первого случая получаем

$$A_{mn} = \frac{\sum_k Q_{mnk}}{\mu}$$

Найдем μ из ограничения вида равенства.

$$1 = \sum_{m,n} A_{mn} = \sum_{m,n} \frac{\sum_k Q_{mnk}}{\mu + \lambda_{mn}}$$

Для ненулевых A_{mn} в сумме по m, n имеем $\lambda_{mn} = 0$.

$$1 = \sum_{m,n:A_{mn}>0} \frac{\sum_k Q_{mnk}}{\mu} \Rightarrow \mu = \sum_k \sum_{m,n:A_{mn}>0} Q_{mnk}$$

$Q_{mnk} = 0$ для таких $A_{mn} = 0$, это следует из выражения в E-шаге, то есть на суммирование в 1 не влияют. Так как $q(d_k)$ - это распределение, то сумма по всем возможным m, n даст 1.

$$\mu = \sum_k \sum_{m,n:A_{mn}>0} Q_{mnk} = K$$

$$\boxed{A_{mn} = \frac{\sum_k Q_{mnk}}{K}}$$

В случае MAP-оценки на $q(d_k)$ получаем

$$\boxed{A_{mn} = \frac{\sum_k \mathbb{I}[m = m^k_{MAP}, n = n^k_{MAP}]}{K}}$$

где m^k_{MAP}, n^k_{MAP} — сдвиги, для которых достигается максимум апостериорного распределения.

Поиск оптимального F

$$L = \sum_{m,n,k} Q_{mnk} \log p(X_k \mid d_k = (m,n), \theta) + const =$$

$$= \sum_{m,n,k} Q_{mnk}(\sum_{(i,j)\in FA(m,n)} -\frac{1}{2s^2}(X_k(i,j) - F(i-m, j-n))^2) + const$$

$$\frac{\partial L}{\partial F_{ab}} = \sum_{m,n,k} \frac{Q_{mnk}}{s^2}(X_k(a+m, b+n) - F(a,b)) = 0$$

$$\boxed{F_{ab} = \frac{\sum_{m,n,k} Q_{mnk} X_k(a+m, b+n)}{K}}$$

В случае MAP-оценки на $q(d_k)$ получаем

$$\boxed{F_{ab} = \frac{\sum_{m,n,k} \mathbb{I}[m = m^k_{MAP}, n = n^k_{MAP}] X_k(a+m, b+n)}{K}}$$

Поиск оптимального B

$$L = \sum_{m,n,k} Q_{mnk} \log p(X_k \mid d_k = (m,n), \theta) + const =$$

$$= \sum_{m,n,k} Q_{mnk}(\sum_{(i,j)\notin FA(m,n)} -\frac{1}{2s^2}(X_k(i,j) - B(i,j))^2) + const$$

$$\frac{\partial L}{\partial B_{ij}} = \sum_{m,n,k} \frac{Q_{mnk}}{s^2}(X_k(i,j) - B(i,j)) \mathbb{I}[(i,j) \notin FA(m,n)] = 0$$

$$\boxed{B_{ij} = \frac{\sum_{m,n,k} Q_{mnk} X_k(i,j) \mathbb{I}[(i,j) \notin FA(m,n)]}{\sum_{m,n,k} Q_{mnk} \mathbb{I}[(i,j) \notin FA(m,n)]} = \frac{\sum_k X_k(i,j) - \sum_{m,n,k} Q_{mnk} X_k(i,j) \mathbb{I}[(i,j) \in FA(m,n)]}{K - \sum_{m,n,k} Q_{mnk} \mathbb{I}[(i,j) \in FA(m,n)]}}$$

В случае MAP-оценки на $q(d_k)$ получаем

$$\boxed{B_{ij} = \frac{\sum_k X_k(i,j) - \sum_{m,n,k} \mathbb{I}[m = m^k_{MAP}, n = n^k_{MAP}] X_k(i,j) \mathbb{I}[(i,j) \in FA(m,n)]}{K - \sum_{m,n,k} \mathbb{I}[m = m^k_{MAP}, n = n^k_{MAP}] \mathbb{I}[(i,j) \in FA(m,n)]}}$$

Поиск оптимального $\mathbf{s}^2$

i, j бегут по соответствующим областям, в целях экономии места опустил, из контекста все понятно будет, надеюсь.

$$L = \sum_{m,n,k} Q_{mnk} \left(\sum_{i,j} \left[-\frac{1}{2s^2}(X_k(i,j) - F(i-m, j-n))^2 \right] + \sum_{i,j} \left[-\frac{1}{2s^2}(X_k(i,j) - B(i,j))^2 \right] - \frac{1}{2} HW \log s^2 \right) + const$$

$$\frac{\partial L}{\partial s^2} = \sum_{m,n,k} Q_{mnk} \left(\sum_{i,j} \left[\frac{1}{2(s^2)^2}(X_k(i,j) - F(i-m, j-n))^2 \right] + \sum_{i,j} \left[\frac{1}{2(s^2)^2}(X_k(i,j) - B(i,j))^2 \right] - \frac{1}{2} HW \frac{1}{s^2} \right) = 0$$

Умножим уравнение на $2(s^2)^2$.

$$\boxed{s^2 = \frac{\sum_{m,n,k} Q_{mnk} \left(\sum_{i,j} \left[(X_k(i,j) - F(i-m, j-n))^2 \right] + \sum_{i,j} \left[(X_k(i,j) - B(i,j))^2 \right] \right)}{HWK}}$$

В случае MAP-оценки на $q(d_k)$ получаем

$$\boxed{s^2 = \frac{\sum_{m,n,k} \mathbb{I}[m = m^k_{MAP}, n = n^k_{MAP}] \left(\sum_{i,j} \left[(X_k(i,j) - F(i-m, j-n))^2 \right] + \sum_{i,j} \left[(X_k(i,j) - B(i,j))^2 \right] \right)}{HWK}}$$

Нижняя оценка на неполное правдоподобие

$$\mathcal{L} = \sum_{m,n,k} Q_{mnk} \left(\sum_{i,j} \left[-\frac{1}{2s^2}(X_k(i,j) - F(i-m, j-n))^2 \right] + \sum_{i,j} \left[-\frac{1}{2s^2}(X_k(i,j) - B(i,j))^2 \right] - \frac{1}{2} HW \log s^2 + \log A_{mn} \right) -$$
$$- \sum_{m,n,k} Q_{mnk} \log Q_{mnk}$$

Анализ

Протестируйте полученный EM алгоритм на сгенерированных данных. Сильно ли влияет начальное приближение на параметры на результаты работы? Стоит ли для данной задачи запускать EM алгоритм из разных начальных приближений?

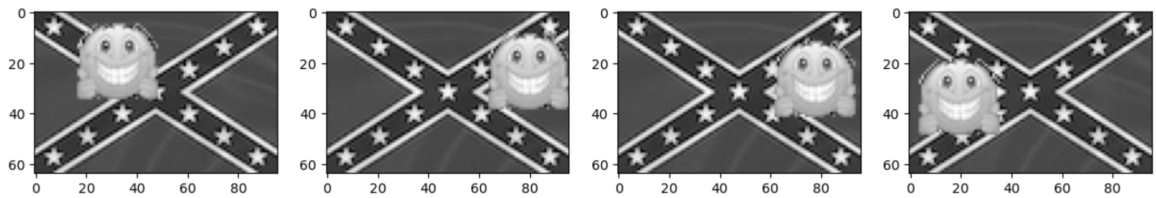

Рис. 1: Пример сгенерированных данных до зашумления

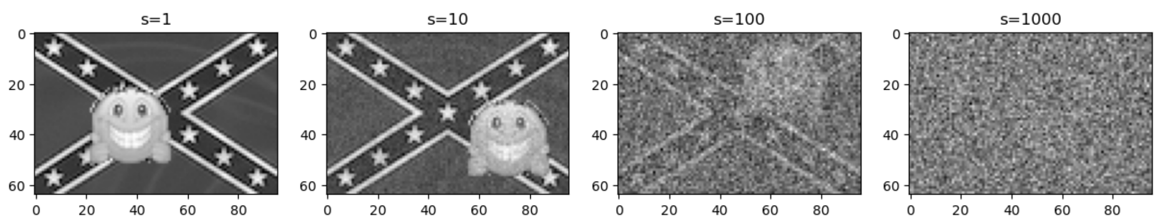

Рис. 2: Пример сгенерированных данных после зашумления с разными значениями s

Действительно, начальное приближение на параметры влияет на результаты, объект в F может быть с разным смещением, но в среднем результаты хорошие и получается покрыть объект почти полностью.

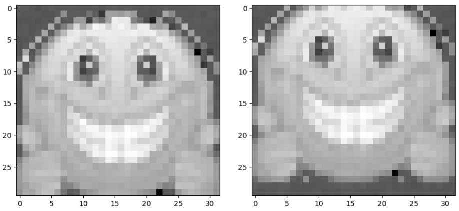

Рис. 3: Результаты запусков из разных инициализаций

Думаю, имеет смысл в данной задаче запускать EM с рестартами для более качественных результатов.

Запустите EM алгоритм на сгенерированных выборках разных размеров и с разным уровнем зашумления. Как изменения в обучающей выборке влияют на результаты работы (получаемые F, B и $\mathcal{L}(q, \theta, A)$)? При каком уровне шума EM-алгоритм перестает выдавать вменяемые результаты? В данном пункте учтите, что для сравнения значения $\mathcal{L}(q, \theta, A)$ для выборок разного размера стоит нормировать его на объем выборки.

Как видим, при увеличении шума получаемое изображение становится все хуже, где-то при s в районе 500 уже нельзя разобрать, что на нем находится. На удивление при росте размера выборки $\mathcal{L}(q, \theta, A)/K$ уменьшается при небольшом шуме. Предполагаю, это может быть связано с тем, что разные позиции F вносят дисперсию и из-за этого правдоподобие уменьшается. При росте шума это значение выравнивается среди разных размеров выборки.

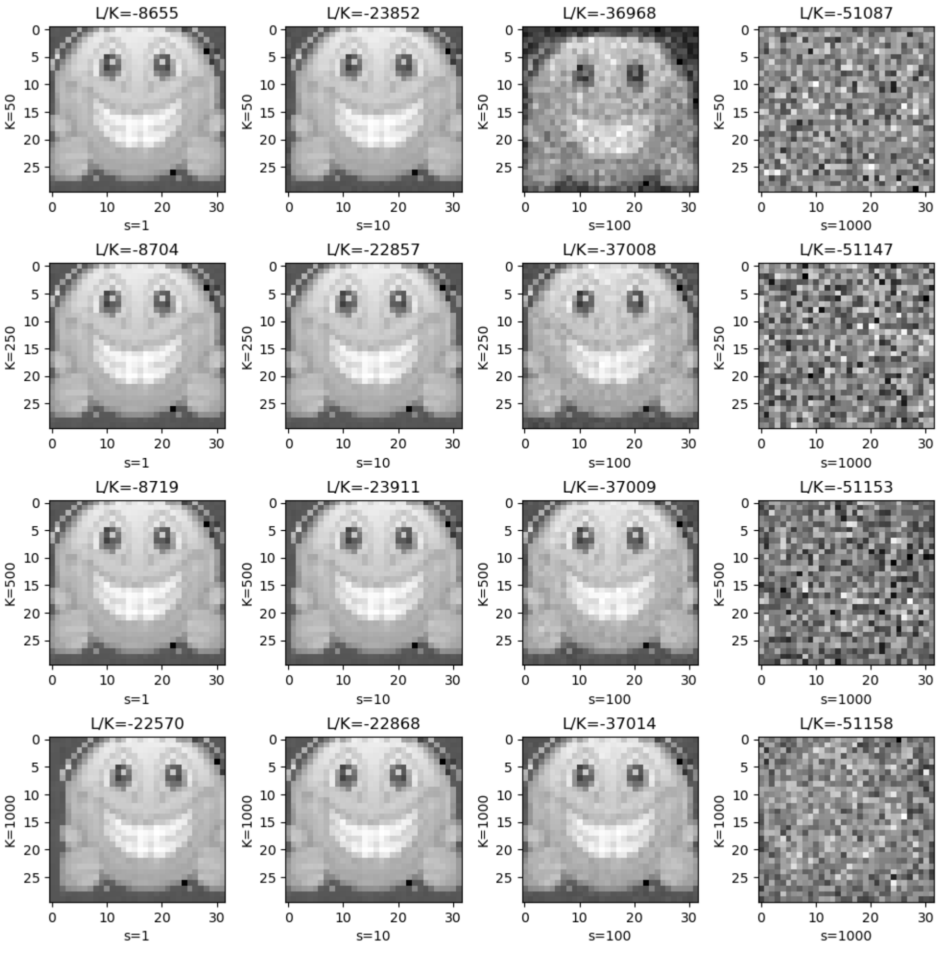

Рис. 4: Восстановленный объект при разных значениях шума и размеров выборки

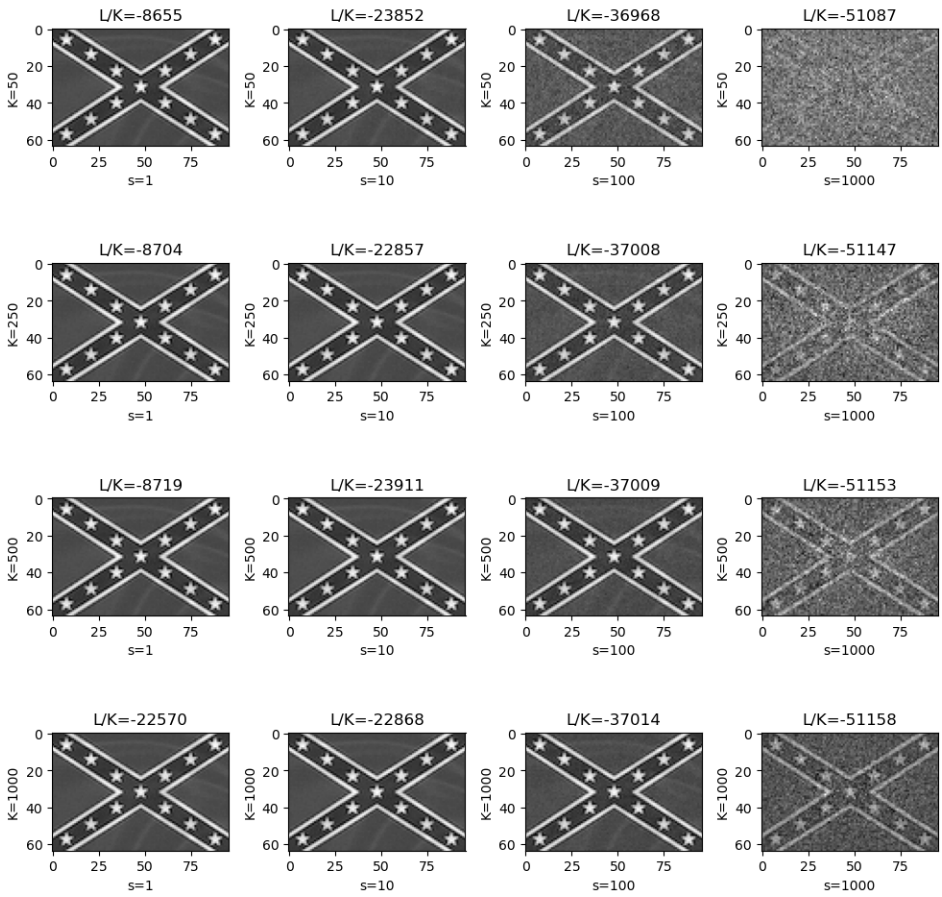

Рис. 5: Восстановленный задний фон при разных значениях шума и размеров выборки

Сравните качество и время работы EM и hard EM на сгенерированных данных. Как Вы думаете, почему разница в результатах работы так заметна?

Эксперимент ставился на 500 изображениях с шумом 250. Как видим, hard EM работает почти в 5 раз быстрее. Очевидно, связано с тем, что в hard EM вычисления не проводятся с полным апостериорным распределением. Однако, мода распределения может быть нерепрезентативна, что мы и видим на картинке ниже, EM справился намного лучше.

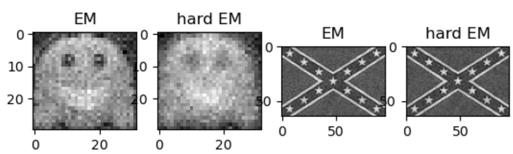

Рис. 6: Результаты работы EM и hard EM

Время работы EM - 2min 35s ± 14.2 s per loop (mean ± std. dev. of 7 runs, 1 loop each)
Время работы hard EM - 33.6 s ± 5.47 s per loop (mean ± std. dev. of 7 runs, 1 loop each)

hard EM быстрее сошелся в свой максимум, который оказался меньше, чем у EM.

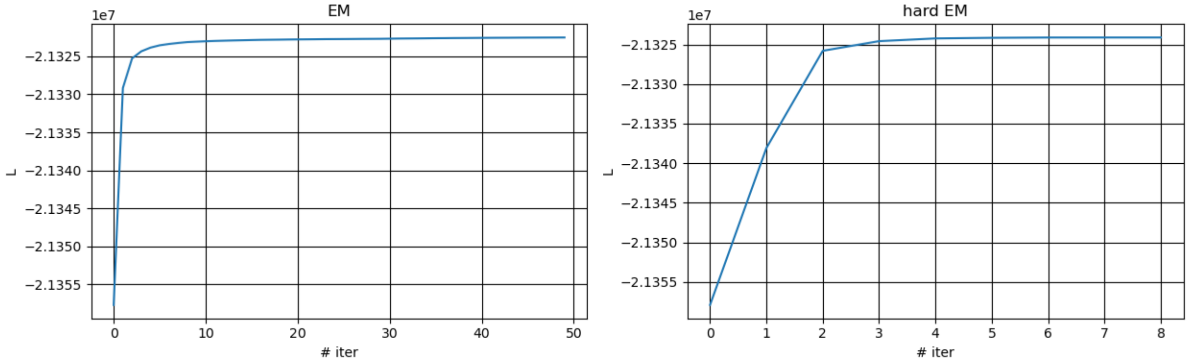

Рис. 7: Сравнение динамики $L(q, \theta, A)$

Примените EM алгоритм к данным с зашумленными снимками преступника. Приведите результаты работы алгоритма на выборках разного размера.

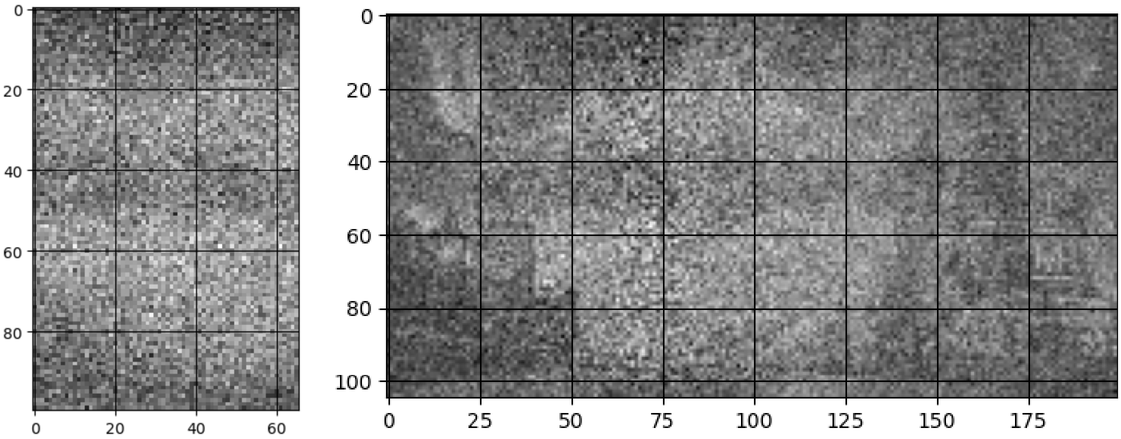

Рис. 8: K=50

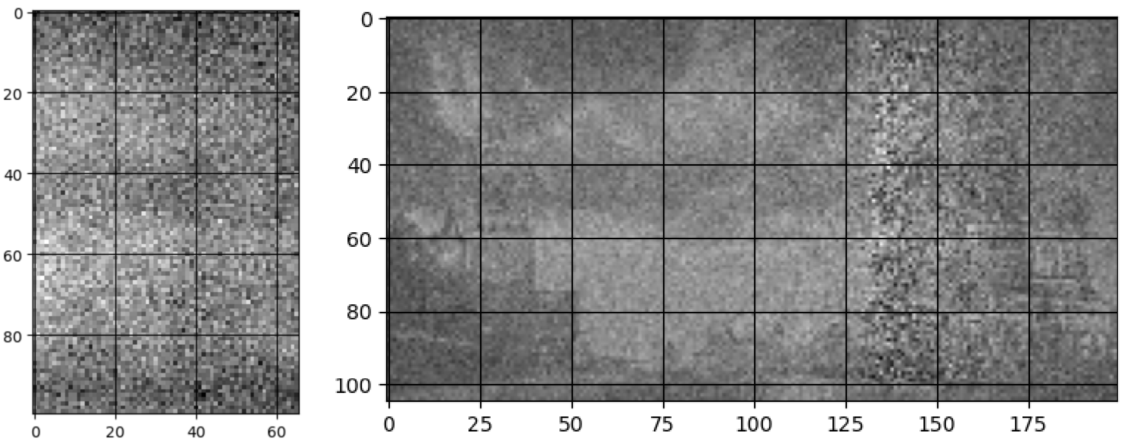

Рис. 9: K=100

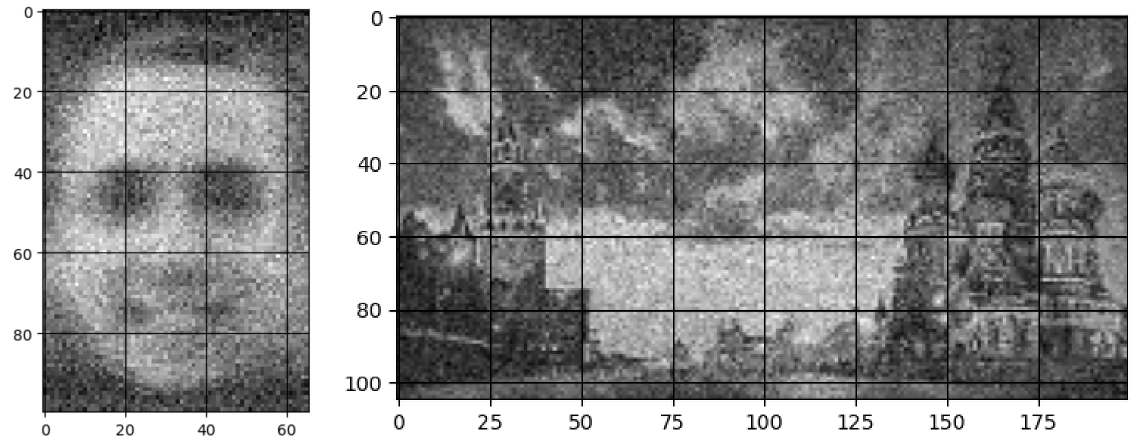

Рис. 10: K=300

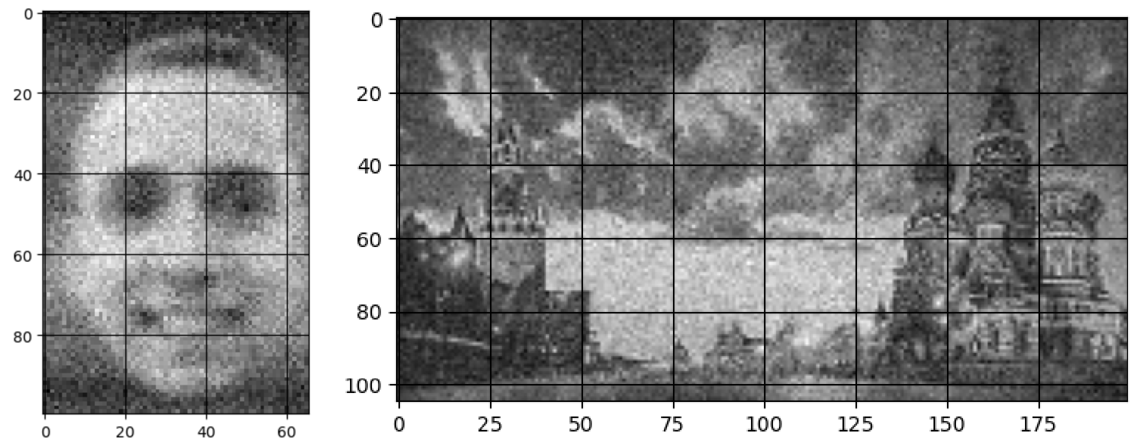

Рис. 11: K=500

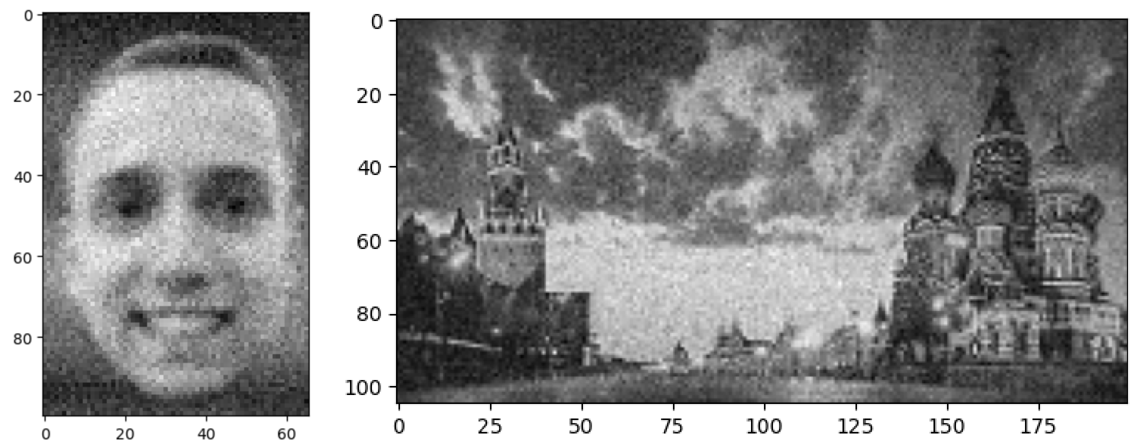

Рис. 12: K=1000

Предложите какую-нибудь модификацию полученного EM алгоритма, которая бы работала на данной задаче качественнее и/или быстрее.

Можно ввести априорное распределение на параметры.

$$F_{ij} \sim \mathcal{N}(F_{ij} \mid 0, \gamma_{ij})$$

$$B_{ij} \sim \mathcal{N}(B_{ij} \mid 0, \delta_{ij})$$

где настоящие картинки живут от -1 до 1. E-шаг не изменится, так как мы максимизируем по q при фиксированных параметрах. В M-шаге добавится $\log p(\theta)$ в ожидании по q.

Также можно применять на итерациях фильтр низких частот для F и B, а именно сворачивать их с гауссовским фильтром, убирая шум.